WARWICK
ECONOMICS

A Semiparametric Network Formation Model with Unobserved Linear Heterogeneity

Luis E. Candelaria

July 2020 No: 1279

Warwick Economics Research Papers

ISSN 2059-4283 (online)
ISSN 0083-7350 (print)

A Semiparametric Network Formation Model with Unobserved Linear Heterogeneity*

Luis E. Candelaria†

July 10, 2020

Abstract

This paper analyzes a semiparametric model of network formation in the presence of unobserved agent-specific heterogeneity. The objective is to identify and estimate the preference parameters associated with homophily on observed attributes when the distributions of the unobserved factors are not parametrically specified. This paper offers two main contributions to the literature on network formation. First, it establishes a new point identification result for the vector of parameters that relies on the existence of a special regressor. The identification proof is constructive and characterizes a closed-form for the parameter of interest. Second, it introduces a simple two-step semiparametric estimator for the vector of parameters with a first-step kernel estimator. The estimator is computationally tractable and can be applied to both dense and sparse networks. Moreover, I show that the estimator is consistent and has a limiting normal distribution as the number of individuals in the network increases. Monte Carlo experiments demonstrate that the estimator performs well in finite samples and in networks with different levels of sparsity.

Keywords: Network formation, Unobserved heterogeneity, Semiparametrics, Special regressor, Inverse weighting.

*First version: November, 2016. A previous version of this paper was titled: "A Semiparametric Network Formation Model with Multiple Linear Fixed Effects."

†Department of Economics, University of Warwick, Coventry, U.K. Email: L.Candelaria@warwick.ac.uk.
I am deeply grateful to Federico Bugni, Shakeeb Khan, Arnaud Maurel, and Matthew Masten for their excellent guidance, constant encouragement, and helpful discussions. I also thank Irene Botosaru, Áureo de Paula, Andreas Dzemski, Cristina Gualdani, Bryan Graham, Bo Honoré, Arthur Lewbel, Thierry Magnac, Chris Muris, James Powell, Adam Rosen, Takuya Ura, Martin Weidner, and seminar participants at Aarhus, Cambridge, Duke, Gothenburg, LSE, Surrey, Syracuse, TSE, UCL, UNC Chapel Hill, Vanderbilt, Warwick, 2018 ES Winter Meeting in Philadelphia, 2019 Panel Data Workshop at the University of Amsterdam, 2019 Royal Economic Society at the University of Warwick, for their comments.

1 Introduction

People tend to connect with individuals whom they share similar observed attributes. This observation is known as homophily and it is one of the main objects of study in the literature of social networks (McPherson, Smith-Lovin, and Cook 2001). Nonetheless, few have investigated the role of homophily when individuals have preferences for unobserved attributes. Proper policy evaluation requires distinguishing among the contribution of these two factors since each has a distinct policy implication. For example, students forming friendships might link based on their similarities on observed socioeconomic attributes as well as on their preferences for high levels of unobserved ability. Whereas the socioeconomic attributes can be influenced by a given policy intervention, preferences for ability are harder to change via targeted policies. In this paper, I study the identification and estimation of the preference parameters associated with the observed attributes in a model of network formation that accounts for valuations on unobserved agent-specific factors. The identification and estimation strategies that I develop do not depend on distributional assumptions of the unobserved random components.

In particular, I consider a semiparametric model of network formation with unobserved agent-specific heterogeneity. Specifically, two distinct agents i and j form an undirected link according to the following network formation equation:[1]

$$D_{ij} = \mathbf{1}\left[g_0(Z_i, Z_j)'\beta_0 + A_i + A_j - U_{ij} \geq 0\right], \tag{1}$$

where $\mathbf{1}[\cdot]$ is the indicator function, D_{ij} is a binary outcome variable that takes a value equal to 1 if agents i and j form a link and 0 otherwise, Z_i is a vector of individual-specific and observed attributes, g_0 is a measurable function that is assumed to be known, nonlinear, finite, and symmetric on its arguments, β_0 is a vector of unknown parameters, A_i and A_j are unobserved and agent-specific random variables, and U_{ij} is an unobserved and link-specific disturbance term.

Intuitively, equation (1) says that an undirected link between two agents is formed if the net benefit of the link between agents i and j is nonnegative. The components in equation (1) can be classified into three different categories. The first class, given by the vector of exogenous attributes $g_0(Z_i, Z_j)$, captures the agents' preferences for establishing a link based on observed characteristics. For instance, this component is known as homophily on observed attributes when it captures preferences for sharing similar traits. The second class, formed by the agent-specific and unobserved factors A_i and A_j, captures the individual preferences for establishing connections based on agent-specific unobserved types. Finally, the third class, given by a link-specific disturbance term U_{ij}, captures the exogenous factors that influence the decision of forming a specific link. The components in the last two categories are known to the agents but unobserved to the researcher.

[1] A link between two agents is undirected if the connection is reciprocal. In other words, two agents are either connected or they are not. It excludes the case that one agent is related to a second one without the second being related to the first.

The agent-specific factors in equation (1) allow for unobserved heterogeneity across the individuals' decisions; this property enables the model to predict network structures with individual connections that are heterogeneous. Moreover, under an unrestricted distribution of the unobserved agent-specific factors, these components could exhibit flexible dependence with the observed attributes.

This paper offers two main contributions to the literature on network formation. The first contribution is to propose a new point identification strategy to identify the vector of coefficients in a semiparametric network formation model with unobserved agent-specific factors. The point identification result is, to the best of my knowledge, the first generalization of a special regressor to analyze a network formation model (Lewbel 1998 and Lewbel 2000). This result depends on the existence of a special regressor and follows from weighting each linking decision in the network by the inverse of the conditional density of the special regressor given the observed attributes. In section 3.1, I provide sufficient conditions to point identify the vector of coefficients. In section 3.2, I provide a second point identification result that does not assume the existence of a special regressor. This result requires that at least one covariate has a full support and consists of finding a sufficient statistic for the unobserved heterogeneity in equation (1) at the tails of the distribution of the observed covariate with full support.

As a second contribution, I use the point identification result in section 3.1 to introduce a two-step semiparametric estimator of the vector of coefficients with a first-stage kernel estimator. As an appealing property, this estimator has a closed-form and is computationally tractable. In section 4, I provide sufficient conditions to show that the estimator is consistent, and it has a limiting normal distribution. I perform inference in a setting when only one network with a large number of agents is observed in the data. Furthermore, I propose an adaptive inference approach to adjust for varying rates of convergence due to different levels of sparsity in the network (see, e.g., Andrews and Schafgans 1998 and Khan and Tamer 2010).

In the rest of this section, I relate my results to the existing literature.

This paper is most closely related to the literature that studies dyadic network formation models with unobserved heterogeneity, (see, e.g., Graham 2017, and Graham 2019a,b for additional surveys). Within this literature, the studies by Charbonneau (2017); Jochmans (2017, 2018); Dzemski (2019), and Yan, Jiang, Fienberg, and Leng (2019) have analyzed the formation of a directed network.[2] Their methodologies differ substantially from the one proposed here since they follow a parametric conditional maximum likelihood approach to estimate the vector of coefficients β_0. In contrast, I study the formation of an undirected network and follow a semiparametric approach.

This paper builds on the seminal work by Graham (2017), which aims to detect preferences

[2] Charbonneau (2017) and Jochmans (2017) study a two-way gravity model, which can be rationalized as a bipartite network with directed links.

for homophily in an undirected network model with agent heterogeneity. Graham (2017, p. 1040) introduces a Tetrad Logit Estimator with identification and asymptotic properties that depend on the link-specific disturbance terms following a logistic distribution. The point identification and estimation results presented below relax this requirement and can be applied to models where the distribution of U_{ij} is not parametrically specified.

Since the initial draft of this paper was circulated, recent studies have appeared analyzing semiparametric or nonparametric variations of a dyadic network formation model with unobserved heterogeneity; these papers include those by Toth (2017); Gao (2020), and Zeleneev (2020).

Similarly to this paper, Toth (2017) study a dyadic network formation model in which the distribution of U_{ij} is unknown. However, the author uses a different identification strategy. In particular, his strategy relies on assuming that each component in the vector of observed attributes Z_i is continuously distributed which is then used to propose an identification strategy similar to the maximum rank by Han (1987). An estimator for β_0 is then defined as the maximizer of a U process of order 4, with a nonparametric first-step estimator.[3]

Gao (2020) studies the identification of a dyadic network model with a nonparametric functional form for the preferences on homophily and an unknown cumulative distribution for U_{ij}.[4] He identifies the nonparametric homophily function by introducing a novel identification strategy that imposes as stochastic restrictions on the distribution of U_{ij}, an interquartile-range normalization and a location normalization of one of the quantiles.

Finally, Zeleneev (2020) studies the identification and estimation of a dyadic network formation model with a nonparametric structure of the unobserved heterogeneity. This framework allows him to account for latent homophily on the unobserved attributes. The author identification analysis is based on introducing a pseudo-distance between a pair of agents i and j, which allows him to recover groups of agents with the same levels of agent-specific unobserved heterogeneity. After conditioning on the matched agents with similar unobserved heterogeneity, the identification of the vector of coefficients proceeds from a pairwise difference strategy. The estimation procedure follows the same logic of the identification strategy.

Contrary to the previous studies, the identification strategy proposed here is based on the existence of a special regressor, (see, e.g., Lewbel (1998) and Lewbel (2012) for a survey). This paper, to the best of my knowledge, represents the first effort in the econometric literature to introduce a special regressor to analyze a network formation model. The vector of parameters β_0 is point identified after introducing a transformation that consists of weighting the linking decisions

[3] Toth (2017) also proposes a variation of his estimation strategy which requires maximizing a U-process of order 2, with a nonparametric first-step estimator. This moditication improves the computational tractability of his method.

[4] Gao (2020) also provides several interesting extensions on the functional form of the unobserved heterogeneity; for reference, see Gao (2020, p. 5) and Zeleneev (2020, p. 6). Those extensions are beyond the scope of this paper and left for future research.

D_{ij} by the inverse of the conditional density of the special regressor given the observed attributes. This transformation utilizes features of the distributions of observables and does not represent a stochastic restriction on the distribution of U_{ij}. Therefore it is not nested in any existing work. As a restriction on the distribution of U_{ij}, I normalize to zero the conditional mean of the link-specific disturbance terms given the observed attributes.[5] In Section 3.1, I provide a detailed discussion on the sufficient conditions needed to point identify β_0 via the existence of a special regressor.

The second point identification result introduced in section 3.2 is based on a sufficient statistic argument at the tails of the distribution of a covariate with full support. The identification strategy shows that within- and across-individuals variation in the linking decisions can be used as a sufficient statistic to differentiate out the unobserved agent-specific factors in some sets of sufficient variations of the covariate with full support. The existence of only one continuous attribute with large support in Z_i is sufficient to show this result. The latter assumption is satisfied by many real network datasets (e.g., household income in the Add Health dataset), and hence it is empirically relevant. The resulting semiparametric estimator is solved in one step, and it is defined as the maximizer of a U-process of order 4 with a trimming sequence.

In Section 4, I introduce a two-step semiparametric estimator for β_0 based on the identification result that requires the existence of a special regressor. The estimator has an analytic form similar to the least-squares, and it uses a first-step kernel estimator to weight the linking decisions D_{ij} by the inverse of the conditional density of the special regressor. In a recent paper, Graham, Niu, and Powell (2019) have studied the nonparametric estimation of density functions with dyadic data. I follow their findings to perform the first-step kernel estimation. In theorems 4.1 and 4.2, I show that the semiparametric estimator for β_0 is consistent and has limiting normal distribution.

Finally, the network formation model that I analyze is related to the literature on empirical games. Specifically, the model in equation (1) can be derived as a stable outcome in a static game. Papers that study the strategic formation of a network as a static game include Goldsmith-Pinkham and Imbens (2013); Leung (2015a,b); Menzel (2015); Miyauchi (2016); Boucher and Mourifié (2017); de Paula, Richards-Shubik, and Tamer (2017); Mele (2017); Candelaria and Ura (2018); Sheng (2018); Gualdani (2020), and Ridder and Sheng (2020). The authors study network formation models that account for network externalities. Network externalities generate interdependencies in the linking decisions that depend on the structure of the network. The identification and estimation methods used in these papers differ substantially from the ones proposed here as they restrict the presence of unobserved agent-specific heterogeneity.

The rest of the paper is organized as follows. Section 2 introduces the network formation model and motivates it as a stable outcome of a random utility model with transferable utilities. Section 3 provides the main identification results of the paper. Section 4 introduced the semiparametric

[5] In further research I will explore the informational content of the special regressor in a network formation model given a quantile or median restriction.

estimator and proves the main asymptotic results. Section 5 reports simulation evidence and section 6 concludes. The appendix collects the proofs of various lemmas and theorems.

2 Network formation model

A network is an ordered pair $(\mathcal{N}_n, \boldsymbol{D}_n)$ formed by a set of n agents denoted by $\mathcal{N}_n = \{1, \cdots, n\}$ and a $n \times n$ adjacency matrix $\boldsymbol{D}_n$, which represents the links between the agents in $\mathcal{N}_n$. Let D_{ij} denote the (i, j)th entry of the matrix $\boldsymbol{D}_n$. I assume the network is undirected and unweighted. A network is undirected if the adjacency matrix is symmetric, that is $D_{ij} = D_{ji}$. A network is unweighted if any (i, j)th entry of the adjacency matrix takes one of two values, where the values are normalized to be 0 and 1. In other words, $D_{ij} \in \{0, 1\}$, where $D_{ij} = 1$ if the agents i and j share a link and $D_{ij} = 0$ otherwise. Furthermore, I normalize the value of self-ties to zero, that is, $D_{ii} = 0$ for any agent i.

Example 1 (Friendships network)**.** *A network of best friends is an example of an undirected and unweighted network. Two agents are considered to be best friends if and only if both agents report each other as friends. In this case, $D_{ij} = D_{ji} = 1$. Also, this example rules out the scenario of an agent reporting herself as her best friend.*

Each agent $i \in \mathcal{N}_n$ is endowed with a $K+1$-dimensional vector of observed attributes Z_i and an unobserved scalar component term A_i. Common examples of observed attributes that could explain the formation of a friendships network among high school students are age, gender, ethnicity, religion, and the students' interest in extracurricular activities. The component A_i captures individual i's preferences for establishing a link based on unobserved and agent-specific attributes. The unobserved component U_{ij} captures exogenous stochastic factors that influence the pair-specific decision of establishing a link between agents i and j.

Given the vectors of observed attributes Z_i and Z_j for $i \neq j$, let $\bar{Z}_{ij} = g_0(Z_i, Z_j)$ be a $K+1$-dimensional vector of pair-specific attributes. The function g_0 is assumed to be a known measurable function that is nonlinear and finite. Given the undirected nature of the network, g_0 is assumed to be symmetric on its terms. The specification of g_0 varies according to the empirical application and is chosen by the researcher to capture homophily or heterophily effects. For example, suppose that Z_i is a scalar random variable that represents agent i's gender, then $\bar{Z}_{ij}$ could be defined as $\mathbf{1}[Z_i = Z_j]$ to capture the preferences for homophily. Under this specification, $\bar{Z}_{ij}$ equals 1 if agents i and j share the same gender and 0 otherwise.

The network formation model described in equation (1) can be obtained as a stable outcome of a random utility model with transferable utilities. In particular, let $\bar{u}_{ij}(\bar{Z}_{ij}, A_j, U_{ij})$ denote individual i's latent valuation of establishing a link with j given their shared-observed attributes $\bar{Z}_{ij}$, agent $j's$

unobserved type A_j, and their common unobserved factor U_{ij}. It follows that the joint net benefit of adding the link $\{i, j\}$ to the network $\boldsymbol{D}_n$ is

$$\bar{u}_{ij}(\bar{Z}_{ij}, A_j, U_{ij}) + \bar{u}_{ji}(\bar{Z}_{ij}, A_i, U_{ij}) = \bar{Z}'_{ij}\beta_0 + A_i + A_j - U_{ij}. \tag{2}$$

Notice that the joint net benefit accounts for the preferences based on the observed attributes $\bar{Z}'_{ij}\beta_0$, as well as preferences for association based on agent-specific factors $A_i + A_j$, and for exogenous factors affecting the decision of establishing a link U_{ij}.

Equation (2) implies that two distinct individuals i and j in $\mathcal{N}_n$ only have valuations for their own observed attributes and agent-specific factors. To clarify, in the link formation decision for dyad $\{i, j\}$, the individuals do not account neither for observed or unobserved attributes of other individuals in the network, nor for the general forms of network externalities other than dyad $\{i, j\}$. These effects are known as network externalities (see, e.g., Chandrasekhar and Jackson 2014; Leung 2015b; Mele 2017; Menzel 2015; Badev 2018; Sheng 2018; Ridder and Sheng 2020). Some examples of these effects are preferences for reciprocity, transitive triads, or high out-degree. I leave this extension as future research.

Next, I introduce the definition of stability.

Definition 1 (Stability)**.** *A network $\boldsymbol{D}_n$ is stable with transfers if for any distinct $i, j \in \mathcal{N}_n$:*

1. *for all $D_{ij} = 1$, $\bar{u}_{ij}(\bar{Z}_{ij}, A_j, U_{ij}) + \bar{u}_{ji}(\bar{Z}_{ij}, A_i, U_{ij}) \geq 0$;*
2. *for all $D_{ij} = 0$, $\bar{u}_{ij}(\bar{Z}_{ij}, A_j, U_{ij}) + \bar{u}_{ji}(\bar{Z}_{ij}, A_i, U_{ij}) < 0$.*

Notice that this definition adapts the pairwise stability in Jackson and Wolinsky (1996) to allow for transfer utilities. Intuitively, this condition states that a link within dyad $\{i, j\}$ is established if the net benefit of that connection is nonnegative. For a generalization to nontransferable utilities, see Gao, Li, and Xu (2020).

2.1 Notation

The following notation will be maintained in the rest of the paper. I will assume that the vector of observed covariates $Z_i = (v_i, X'_i)'$ is comprised by a scalar random variable $v_i \in \mathbb{R}$ and K-dimensional random vector $X_i \in \mathbb{R}^K$. Similarly, denote by

$$\bar{Z}_{ij} \quad = \quad \left(g_0(v_i, v_j), g_0(X_i, X_j)'\right)' = (v_{ij}, W'_{ij})'$$

the observed covariates at dyad level, and let $\beta_0 = (1, \theta'_0)'$.

I will denote the distinct profiles of observed attributes for all the agents in the network as $\boldsymbol{Z}_n = \{Z_i : i \in \mathcal{N}_n\}$, $\boldsymbol{v}_n = \{v_i : i \in \mathcal{N}_n\}$, and $\boldsymbol{X}_n = \{X_i : i \in \mathcal{N}_n\}$. Similarly, let $\boldsymbol{A}_n = \{A_i : i \in \mathcal{N}_n\}$ denote the profile of unobserved attributes. Moreover, let $\boldsymbol{Z}_{-ij} = \{Z_k : k \neq i, j\}$, and $\boldsymbol{A}_{-ij} = \{A_k : k \neq i, j\}$ denote the collection of observed and unobserved attributes for all agents in the network other than agents i and j.

The identification and estimation strategies introduced in sections 3 and 4 use the information contained in subnetworks formed by groups of four distinct agents $\{i_1, i_2, j_1, j_2\}$, also known as tetrads. The following notation is used to describe attributes at the tetrad level. Given a network of size n, there is a total of

$$m_n = 4!\binom{n}{4}$$

ordered tetrads with distinct indices $i_1, i_2, j_1, j_2 \in \mathcal{N}_n$. Let σ be a function that maps these tetrads to the index set $\mathcal{N}_{m_n} = \{1, \cdots, m_n\}$. Thus, each tetrad with distinct indices $\{i_1, i_2, j_1, j_2\}$ corresponds to a unique $\sigma(\{i_1, i_2; j_1, j_2\}) \in \mathcal{N}_{m_n}$.

Given any $\sigma(\{i_1, i_2; j_1, j_2\}) \in \mathcal{N}_{m_n}$, let $v_\sigma = \{v_{i_1}, v_{j_1}, v_{i_2}, v_{j_2}\}$, $X_\sigma = \{X_{i_1}, X_{j_1}, X_{i_2}, X_{j_2}\}$, and $A_\sigma = \{A_{i_1}, A_{j_1}, A_{i_2}, A_{j_2}\}$.

Moreover, define the pairwise variations across observed attributes and linking decisions as follows

$$\begin{aligned}
\tilde{v}_\sigma &= \tilde{v}_{i_1 i_2, j_1 j_2} = (v_{i_1 j_1} - v_{i_1 j_2}) - (v_{i_2 j_1} - v_{i_2 j_2}) \\
\tilde{W}_\sigma &= \tilde{W}_{i_1 i_2, j_1 j_2} = (W_{i_1 j_1} - W_{i_1 j_2}) - (W_{i_2 j_1} - W_{i_2 j_2}) \\
\tilde{D}_\sigma &= \tilde{D}_{i_1 i_2, j_1 j_2} = (D_{i_1 j_1} - D_{i_1 j_2}) - (D_{i_2 j_1} - D_{i_2 j_2}).
\end{aligned}$$

Finally, given any fixed tetrad $\sigma(\{i_1, i_2; j_1, j_2\}) \in \mathcal{N}_{m_n}$, let $\omega_{l_1 l_2} = (v_{l_1 l_2}, X_{l_1}, X_{l_2}, A_{l_1}, A_{l_2})$ denote the profile of attributes at dyad-level and $p_n(\omega_{l_1 l_2}) = P[D_{l_1 l_2} = 1 \mid \omega_{l_1 l_2}]$ denote the probability that a link is created for any $(l_1, l_2) \in \{(i_1, j_1), (i_1, j_2), (i_2, j_1), (i_2, j_2)\}$.

3 Identification

This section introduces the main identification results for the semiparametric network formation model with unobserved agent-specific factors. In particular, section 3.1 presents the main point identification result when a special regressor is available. Section 3.2 introduces a second point identification result when a covariate with full support is available.

3.1 Point Identification Result: Special Regressor

Using the notation introduced in section 2, the rest of the paper considers the following representation for the network formation model specified by equation (1). In particular, agents i and j in $\mathcal{N}_n$ with $i \neq j$ will form an undirected link according to the next equation

$$D_{ij} = \mathbf{1}\left[v_{ij} + W_{ij}'\theta_0 + A_i + A_j - U_{ij} \geq 0\right], \tag{3}$$

where the coefficient associated with v_{ij} has been normalized to 1 and θ_0 is a K-dimensional vector of coefficients. Given that the network of interest is undirected, U_{ij} is assumed to be symmetric, i.e., $U_{ij} = U_{ji}$. The vector θ_0 represents the main parameter of interest.

Assumptions 3.1.1-3.1.5 will specify the underlying structure for the network formation model in equation (3), which will be used to show the main identification result for θ_0.

Assumption 3.1.1. *The random sequence $\{Z_i, A_i\}_{i=1}^n$ is independent and identically distributed.*

Assumption 3.1.1 describes the sampling process, and it is widely used to describe network data, (see, e.g., Graham 2017; Jochmans 2018, and Auerbach 2019).

Assumption 3.1.2. *For any finite n, the following holds.*

1. *The sequence $\{U_{ij} \mid \boldsymbol{Z}_n, \boldsymbol{A}_n\}_{i\neq j}$ is conditionally independent and identically distributed for any dyad $\{i,j\}$. Moreover, $U_{ij} = U_{ji}$ for any dyad $\{i,j\}$.*
2. *For any dyad $\{i,j\}$, $U_{ij} \mid \boldsymbol{Z}_n, \boldsymbol{A}_n \overset{d}{=} U_{ij} \mid Z_i, Z_j, A_i, A_j$.*

Assumption 3.1.2.1 states that conditional on $(\boldsymbol{Z}_n, \boldsymbol{A}_n)$ the link-specific disturbance terms $\{U_{ij}\}_{i\neq j}$ are independent across dyads $\{i,j\}$ and drawn from the same distribution. Furthermore, Assumption 3.1.2.2 requires that conditional on (Z_i, Z_j, A_i, A_j), the link-specific disturbance terms U_{ij} are independent of any observed or unobserved feature in $(\boldsymbol{Z}_{-ij}, \boldsymbol{A}_{-ij})$. Assumption 3.1.2 ensures that each of the linking decisions in the network is conditionally independent. In other words, it rules out interdependence across linking decisions due to externalities across the network.

Notice that Assumption 3.1.2 allows for heteroskedasticity of a general form in the distribution of U_{ij}. Moreover, it allows for a flexible dependence between the unobserved agent-specific factors and the observed attributes. In other words, Assumption 3.1.2 does not restrict the joint distribution $(\boldsymbol{Z}_n, \boldsymbol{A}_n)$. Assumption 3.1.2 is commonly used in semiparametric nonlinear panel data models, for example in Arellano and Honoré (2001). In network formation models, full stochastic independence $U_{ij} \perp \boldsymbol{Z}_n, \boldsymbol{A}_n$ is usually imposed (see, e.g., Leung 2015b; Menzel 2015; Graham 2017; Toth 2017, and Gao 2020). Arbitrary heteroskedasticity is also considered in Zeleneev (2020).

Assumption 3.1.3. *Given n and any distinct $i, j \in \mathcal{N}_n$, let $e_{ij} = A_i + A_j - U_{ij}$ and suppose that e_{ij} is conditional independent of v_{ij} given (X_i, X_j). Let $F_{e|x}(e_{ij} \mid X_i, X_j)$ denote the conditional distribution of e_{ij} given (X_i, X_j), with support given by $\mathbb{S}_e(X_i, X_j)$ and finite first moment.*

Assumption 3.1.3 represents an exclusion restriction, and it entails that the regressor v_{ij} is conditionally independent of e_{ij} given the observed attributes (X_i, X_j).[6] In other words, v_{ij} is a special regressor in the sense of Lewbel (1998), Lewbel (2000), and Lewbel (2012).

Assumption 3.1.4. *Given n and any distinct $i, j \in \mathcal{N}_n$, the conditional distribution of v_{ij} given (X_i, X_j) is absolutely continuous with respect to the Lebesgue measure with conditional density $f_{v|x}(v_{ij} \mid X_i, X_j)$ and support given by $\mathbb{S}_v(X_i, X_j) = [\underline{s}_v, \overline{s}_v]$ for some constants $\underline{s}_v$ and $\overline{s}_v$, with $-\infty \leq \underline{s}_v < 0 < \overline{s}_v \leq \infty$. For any (X_i, X_j), the support of $-W_{ij}'\theta_0 - e_{ij}$ is a subset of the interval $[\underline{s}_v, \overline{s}_v]$.*

Assumption 3.1.4 is a support condition, and it ensures that $v_{ij} \mid X_i, X_j$ has a positive density function $f_{v|x}(v_{ij} \mid X_i, X_j)$ on $\mathbb{S}_v(X_i, X_j)$. Furthermore, it requires that for any (X_i, X_j) the support of $(-W_{ij}'\theta_0 - e_{ij})$ is contained in $\mathbb{S}_v(X_i, X_j)$. Notice that Assumption 3.1.4 does not restrict $v_{ij} \mid X_i, X_j$ to have full support on the real line. Hence the point identification result introduced in this section is general enough to include both cases: (i) the full support case, and (ii) the existence of a continuous covariate with bounded support that contains *supp* $\left(-W_{ij}'\theta_0 - e_{ij} \mid X_i, X_j\right)$. Moreover, observe that Assumption 3.1.4 leaves unrestricted the distribution of the observed attributes (X_i, X_j). Hence, this identification strategy also allows for discrete covariates in W_{ij}.

Assumption 3.1.5. *Given n and any tetrad $\sigma \in \mathcal{N}_{m_n}$, $\mathbb{E}\left[U_{ij} \mid X_i, X_j\right] = 0$, and*

$$\Gamma_0 = \mathbb{E}\left[\tilde{W}_\sigma \tilde{W}_\sigma'\right]$$

is a finite and nonsingular matrix.

The first part of assumption 3.1.5 represents a stochastic restriction on the link-specific disturbance term. In particular, it requires that $U_{ij} \mid X_i, X_j$ has conditionally mean zero. The second part of assumption 3.1.5 is the standard full rank condition on the pairwise variation of the observed attributes $\tilde{W}_\sigma$, and it ensures that θ_0 is point identified.

The network formation model specified by equation (3) and Assumptions 3.1.1-3.1.5 represents, to the best of my knowledge, the first generalization of the special regressor to analyze network

[6] The conditional independence property needs to hold after conditioning on the observed attributes (X_i, X_j), and not just the dyad-specific covariates W_{ij}. The intuition behind this insight follows from Assumption 3.1.1, which allows for unrestricted dependence between X_i, and A_i. In particular, the proof of Theorem 3.1 requires that any stochastic variation left in $A_i + A_j$ after conditioning on (X_i, X_j), is independent of W_{kl} for any $k, l \in \mathcal{N}_n$, including, for example W_{il}. This property no longer holds if the conditioning variable used is W_{ij} since it is only a feature of (X_i, X_j).

data. Following Lewbel (1998, 2000), Honoré and Lewbel (2002), and Chen, Khan, and Tang (2019), let D^*_{ij} be defined as

$$D^*_{ij} = \left[\frac{D_{ij} - \mathbf{1}[v_{ij} > 0]}{f_{v|x}(v_{ij} \mid X_i, X_j)}\right] \tag{4}$$

for any distinct $i, j \in \mathcal{N}_n$.

The following theorem and appended corollary formalize the first point identification result for θ_0.

Theorem 3.1. *If Assumptions 3.1.3-3.1.5 hold in equation (3), then for any distinct i and j in $\mathcal{N}_n$*

$$\mathbb{E}[D^*_{ij} \mid X_i, X_j] = W'_{ij}\theta_0 + \mathbb{E}[A_i + A_j \mid X_i, X_j].$$

Proof. See Appendix A. □

Corollary 3.1. *If Assumptions 3.1.1-3.1.5 hold in equation (3), then for any tetrad $\sigma \in \mathcal{N}_{m_n}$*

$$\mathbb{E}\left[\tilde{W}_\sigma \tilde{D}^*_\sigma\right] = \mathbb{E}\left[\tilde{W}_\sigma \tilde{W}'_\sigma\right]\theta_0, \tag{5}$$

and hence,

$$\theta_0 = \Gamma_0^{-1} \times \Psi_0 \tag{6}$$

with $\Psi_0 = \mathbb{E}\left[\tilde{W}_\sigma \tilde{D}^*_\sigma\right]$.

Proof. See Appendix A. □

Theorem 3.1 and Corollary 3.1 demonstrate that θ_0 is point identified using the information contained in the joint distribution of $\{\tilde{D}^*_\sigma, \tilde{W}_\sigma\}$ at tetrad level, and with analytic expression given by equation (6). This result shows that θ_0 is identified as an average of the linking decisions $\tilde{D}_\sigma$ which are weighted by the inverse of the conditional density of the special regressor given the observed attributes, $f_{v|x}(v_{ij} \mid X_i, X_j)$. The result in Corollary 3.1 will be used as a foundation of the semiparametric estimator introduced in Section 4.

Given the results in Theorem 3.1 and Corollary 3.1 the average contribution of the unobserved agent-specific factors to the formation of a link is also identified.

Corollary 3.2. *If Assumptions 3.1.1-3.1.5 hold in equation (3), then for any i and j in $\mathcal{N}_n$*

$$\mathbb{E}[A_i + A_j] = \mathbb{E}\left[D^*_{ij}\right] - \mathbb{E}[W_{ij}]'\theta_0, \tag{7}$$

3.2 Second Point Identification Result

In this section, I provide a second point identification result for the vector of coefficients θ_0. This result does not require the regressor v_{ij} to be conditionally independent of the unobserved terms, $A_i + A_j - U_{ij}$. Nonetheless, it imposes a large support condition on v_{ij} and bounds the contribution that the unobserved heterogeneity $A_i + A_j$ has on the formation of links.

The following notation will be used to state and proof this result. For any fixed tetrad $\sigma(\{i, j; k, l\}) \in \mathcal{N}_{m_n}$, denote the profile of observed attributes at tetrad level as $\bar{\boldsymbol{v}}_\sigma = (v_{ik}, v_{il}, v_{jk}, v_{jl})$ and $\bar{\boldsymbol{Z}}_\sigma = (\bar{\boldsymbol{v}}_\sigma, X_\sigma)$. Moreover, for any $\sigma(\{i, j; k, l\}) \in \mathcal{N}_{m_n}$ and agent r with $r \in \{i, j\}$ denote the within-individual r variation of the observed attributes as $\Delta_\sigma v_r = v_{rk} - v_{rl}$ and $\Delta_\sigma W_r = W_{rk} - W_{rl}$, and the within-individual r variation of the unobserved attributes as $\Delta_\sigma A = A_k - A_l$.

The following assumptions are sufficient to show this result.

Assumption 3.2.1. *For any finite n and dyad $\{i, j\}$, Assumption 3.1.2 holds. Furthermore, the link-specific unobserved term $U_{ij} \mid Z_i, Z_j, A_i, A_j$ has a positive density over the real line.*

Assumption 3.2.1 ensures that the disturbance term U_{ij} has a large support for any value of (Z_i, Z_j, A_i, A_j). This assumption is used for simplicity to ensure that the conditional probability of forming a link is well defined for any value of (Z_i, Z_j, A_i, A_j). Notice that any model where the disturbance term U_{ij} is logistically or normally distributed will satisfy this condition.

Assumption 3.2.2. *The parameter space Θ is compact.*

Assumption 3.2.2 is a standard assumption in the semiparametrics literature, (see, e.g., Manski 1975, 1985; Newey and McFadden 1994, and Powell 1994). This assumption is used to control the contribution that the variation on W_{ij} has on the formation of links.

Assumption 3.2.3. *For any finite n, the following holds for any $\sigma(\{i, j; k, l\}) \in \mathcal{N}_{m_n}$.*

1. *For all X_σ, $\bar{\boldsymbol{v}}_\sigma$ is continuously distributed with a positive density over $\mathbb{R}^4$.*
2. *For all X_σ and $r \in \{i, j\}$, $\Delta_\sigma v_r$ is continuously distributed with a positive density over the real line, and the* $\operatorname{supp}(-\Delta_\sigma W_r'\theta_0 - \Delta_\sigma A \mid X_\sigma) = [\underline{s}_\varepsilon, \overline{s}_\varepsilon]$ *is known with $-\infty < \underline{s}_\varepsilon < 0 < \overline{s}_\varepsilon < \infty$.*

Assumption 3.2.3 ensures that the regressor v_{ij} has a large support. Moreover, it requires that the variation in v_{ij} dominates the contribution that the remaining factors have in creating a network link. Notice that this condition does not impose that v_{ij} is conditionally independent of $A_i + A_j$ given X_σ. Intuitively, Assumption 3.2.3 guarantees that the information at the tails of the distribution of $\Delta_\sigma v_r$ can disentangle the contributions of the preferences for homophily and unobserved heterogeneity on the creation of network links.

Assumption 3.2.4. *For any finite n and tetrad $\sigma(\{i,j;k,l\}) \in \mathcal{N}_{m_n}$, $P\left[\tilde{W}_\sigma'\gamma \neq 0\right] > 0$ for all non-zero vectors $\gamma \in \mathbb{R}^K$.*

Assumption 3.2.4 is a rull rank condition.

For any fixed $\sigma(\{i,j;k,l\}) \in \mathcal{N}_{m_n}$ and given X_σ, let $\mathcal{V}(X_\sigma)$ denote the set of values for which the variations in $\Delta_\sigma v_i$ and $\Delta_\sigma v_j$ dominates the contribution of the remaining factors. That is

$$\mathcal{V}(X_\sigma) \quad = \quad \{\bar{\boldsymbol{v}}_\sigma : \Delta_\sigma v_i \leq \underline{s}_\varepsilon \ \& \ \Delta_\sigma v_j \geq \overline{s}_\varepsilon, \quad \text{or} \quad \Delta_\sigma v_i \geq \overline{s}_\varepsilon \ \& \ \Delta_\sigma v_j \leq \underline{s}_\varepsilon\}. \tag{8}$$

Notice that this set can be characterized using Assumption 3.2.3. Also, define $\xi(\theta)$ as

$$\xi(\theta) \quad = \quad \left\{ \bar{\boldsymbol{z}}_\sigma : \bar{\boldsymbol{v}}_\sigma \in \mathcal{V}(X_\sigma) \quad \text{and} \quad \begin{array}{c} \text{sign}\left\{\mathbb{E}_{\theta_0}\left[\tilde{D}_\sigma \mid X_\sigma, \bar{\boldsymbol{v}}_\sigma \in \mathcal{V}(X_\sigma), \tilde{D}_\sigma \in \{-2,2\}\right]\right\} \\ \neq \text{sign}\left\{\mathbb{E}_{\theta}\left[\tilde{D}_\sigma \mid X_\sigma, \bar{\boldsymbol{v}}_\sigma \in \mathcal{V}(X_\sigma), \tilde{D}_\sigma \in \{-2,2\}\right]\right\} \end{array} \right\},$$

which characterizes the set of states for which the sign of the conditional expectation of the pairwise variations of the links $\tilde{D}_\sigma$ implied by θ differs from the sign of the conditional expectation generated under θ_0. In other words, the set $\xi(\theta)$ summarizes the values of observed attributes for which θ can be identified from θ_0 using the information contained in the conditional expectation of $\tilde{D}_\sigma$. Hence, θ_0 is said to be identified relative to $\theta \neq \theta_0$ if

$$P\left[\bar{\boldsymbol{Z}}_\sigma \in \xi(\theta)\right] > 0.$$

The next theorem and appended corollary formalizes the second point identification result.

Theorem 3.2. *Suppose Assumptions 3.1.1, 3.2.1, 3.2.2, and 3.2.3 hold in equation (3). Let*

$$Q_\theta \quad = \quad \left\{\bar{z}_\sigma : \bar{\boldsymbol{v}}_\sigma \in \mathcal{V}(X_\sigma) \quad \textit{and} \quad \tilde{W}_\sigma'\theta_0 \leq -\tilde{v}_\sigma < \tilde{W}_\sigma'\theta \quad \textit{or} \quad \tilde{W}_\sigma'\theta \leq -\tilde{v}_\sigma < \tilde{W}_\sigma'\theta_0\right\}.$$

If $P\left[\bar{\boldsymbol{Z}}_\sigma \in Q_\theta\right] > 0$, θ_0 is point identified relative to θ.

Proof. See Appendix A. □

Corollary 3.3. *Suppose Assumptions 3.1.1, 3.2.1- 3.2.4 hold in equation (3). Then θ_0 is point identified.*

Proof. See Appendix A. □

The results in Theorem 3.2 and Corollary 3.3 can be used to define an estimator for θ_0 as the maximizer of a U-process of order 4 with a trimming sequence γ_n such that $\gamma_n \to \infty$ as $n \to \infty$.

In particular, the estimator of θ_0 can be defined as

$$\hat{\theta} \quad = \quad \arg\max_{\theta\in\Theta} \hat{H}_n(\theta, \gamma_n)$$

where

$$\begin{aligned}
\hat{H}_n(\theta, \gamma_n) &= \left[4!\binom{n}{4}\right]^{-1} \sum_{i_1=1}^{n} \sum_{j_1\neq i_1} \sum_{i_2\neq i_1, j_1} \sum_{j_2\neq i_1, j_1, i_2} H\left(\bar{\boldsymbol{Z}}_{\sigma(\{i_1,j_1;i_2,j_2\})}, \tilde{D}_{\sigma(\{i_1,j_1;i_2,j_2\})}; \theta, \gamma_n\right) \\
H\left(\bar{\boldsymbol{Z}}_{\sigma}, \tilde{D}_{\sigma}; \theta, \gamma_n\right) &= \left[\text{sign}\left\{\tilde{v}_\sigma + \tilde{W}_\sigma'\theta\right\} \times \tilde{D}_\sigma\right] \times \mathbf{1}\left[\mid \tilde{D}_\sigma \mid = 2\right] \times \mathbf{1}\left[\mid \Delta_\sigma v_i \mid, \mid \Delta_\sigma v_j \mid \geq \gamma_n\right].
\end{aligned}$$

Although point identification of θ_0 is achieved assuming that the bounds $[\underline{s}_\varepsilon, \overline{s}_\varepsilon]$ are known, notice that they are not needed to define the estimator $\hat{\theta}$. In other words, it is sufficient to assume that $\Delta_\sigma v_i$ has a large support which contains $\text{supp}\left(-\Delta_\sigma W_i'\theta_0 - \Delta_\sigma A \mid X_\sigma\right)$ to characterize the estimator for θ_0.

Naturally, the asymptotic properties of $\hat{\theta}$ will depend on the frequency of subgraph configurations that satisfy the restriction $\mathbf{1}\left[\mid \tilde{D}_\sigma \mid = 2\right]$ in the sample, and the rate at which $\gamma_n \to \infty$ as $n \to \infty$. The rest of this paper prioritizes the study of the semiparametric estimator introduced in section 4 since it is computationally more tractable than $\hat{\theta}$.

4 Inference

In this section, I introduce a semiparametric estimator for θ_0 based on the point identification result derived in section 3.1. The estimator for θ_0 denoted by $\widehat{\theta}_n$ is a two-step estimator with a nonparametric estimate of the conditional distribution of v_{ij} given $\{X_i, X_j\}$, i.e., $f_{v|x}(v_{ij} \mid X_i, X_j)$. Section 4.1 provides sufficient conditions to study the large sample properties of $\widehat{\theta}_n$. Theorem 4.1 proves that $\widehat{\theta}_n$ is a consistent estimator of θ_0. Theorem 4.2 shows that the limiting distribution of $\widehat{\theta}_n$ is normal.

4.1 Consistency

The estimator for θ_0 is defined as the sample analog of equation (6) and is obtained by averaging over the linking decisions $\tilde{D}_\sigma$ for all distinct tetrads $\sigma \in \mathcal{N}_{m_n}$. Given that the inverse of $f_{v|x}(v_{ij} \mid X_i, X_j)$ is used as a weight in the definition of Ψ_0, and hence θ_0, I introduce a trimming sequence intended to avoid boundary effects arising from the first-step estimation of $f_{v|x}(v_{ij} \mid X_i, X_j)$.

Recall that $\tilde{D}_\sigma$ is defined as the pairwise variation across the linking decisions for a given tetrad $\sigma\left(\{i_1, i_2; j_1, j_2\}\right) \in \mathcal{N}_{m_n}$. I extend that notation to define as follows the pairwise variation of the trimmed network links given a trimming parameter τ

$$\begin{aligned}\widetilde{D}^*_{\sigma,\tau} &= \left(D^*_{i_1j_1,\tau} - D^*_{i_1j_2,\tau}\right) - \left(D^*_{i_2j_1,\tau} - D^*_{i_2j_2,\tau}\right)\\ \widehat{D}^*_{\sigma,\tau} &= \left(\widehat{D}^*_{i_1j_1,\tau} - \widehat{D}^*_{i_1j_2,\tau}\right) - \left(\widehat{D}^*_{i_2j_1,\tau} - \widehat{D}^*_{i_2j_2,\tau}\right),\end{aligned}$$

where for any distinct i_1 and j_1 in $\mathcal{N}_n$

$$\begin{aligned}D^*_{i_1j_1,\tau} &= \left(\frac{D_{i_1j_1} - \mathbf{1}\left[v_{i_1j_1} > 0\right]}{f_{v|x}(v_{i_1j_1} \mid X_{i_1}, X_{j_1})}\right) I_\tau(v_{i_1j_1}, X_{i_1}, X_{j_1})\\ \widehat{D}^*_{i_1j_1,\tau} &= \left(\frac{D_{i_1j_1} - \mathbf{1}\left[v_{i_1j_1} > 0\right]}{\widehat{f}_{v|x}(v_{i_1j_1} \mid X_{i_1}, X_{j_1})}\right) I_\tau(v_{i_1j_1}, X_{i_1}, X_{j_1}).\end{aligned}$$

In the equations above, $f_{v|x}(v_{i_1j_1} \mid X_{i_1}, X_{j_1})$ denotes the true conditional density function of $v_{i_1j_1}$ given (X_{i_1}, X_{j_1}), and $\widehat{f}_{v|x}(v_{i_1j_1} \mid X_{i_1}, X_{j_1})$ denotes a kernel estimator of the conditional density of $v_{i_1j_1}$ given (X_{i_1}, X_{j_1}). Thus, $\widetilde{D}^*_{\sigma,\tau}$ denotes the pairwise variation of the trimmed network links assuming that the conditional distribution of the special regressor given the observed attributes is known. Conversely, $\widehat{D}^*_{\sigma,\tau}$ denotes the pairwise variation of the trimmed network links when $f_{v|x}(v_{i_1j_1} \mid X_{i_1}, X_{j_1})$ is replaced by a first-stage kernel estimator $\widehat{f}_{v|x}(v_{i_1j_1} \mid X_{i_1}, X_{j_1})$

The trimming sequence $I_\tau(v_{i_{1j_1}}, X_{i_1}, X_{j_1})$ is a function of the observed attributes at a dyad level, and it converges to 1 as the trimming parameter $\tau \to 0$ when $n \to \infty$. Assumptions 4.1.2 and 4.1.5 below describe the conditions imposed on the trimming parameter τ, (cf., Honoré and Lewbel 2002 and Khan and Tamer 2010).

To ease the exposition, I introduce the following notation for any distinct $i_1, j_1 \in \mathcal{N}_n$

$$\begin{aligned}I_{\tau,i_1j_1} &= I_\tau(v_{i_1j_1}, X_{i_1}, X_{j_1})\\ f_{vx,i_1j_1} &= f_{v,x}(v_{i_1j_1}, X_{i_1}, X_{j_1})\\ f_{x,i_1j_1} &= f_x(X_{i_1}, X_{j_1})\\ \varphi_{i_1j_1} &= D_{i_1j_1} - \mathbf{1}\left[v_{i_1j_1} > 0\right]\\ \varphi_{i_1j_1,\tau} &= \varphi_{i_1j_1} I_{\tau,i_1j_1}.\end{aligned}$$

With this notation at hand, the semiparametric estimator for θ_0 is defined as

$$\widehat{\theta}_n = \widehat{\Gamma}_n^{-1} \times \widehat{\Psi}_{n,\tau} \tag{9}$$

where

$$
\begin{aligned}
\widehat{\Gamma}_n &= \frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} \left[\tilde{W}_\sigma \tilde{W}'_\sigma \right] \\
\widehat{\Psi}_{n,\tau} &= \frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} \left[\tilde{W}_\sigma \widehat{D}^*_{\sigma,\tau} \right]
\end{aligned}
$$

and $m_n = 4!\binom{n}{4}$.

The first-stage kernel estimator $\widehat{f}_{v|x}(v_{i_1 j_1} \mid X_{i_1}, X_{j_1})$ is defined as the ratio of the kernel estimators $\widehat{f}_{vx,i_1 j_1}$ and $\widehat{f}_{x,i_1 j_1}$ with

$$
\begin{aligned}
\widehat{f}_{vx,i_1 j_1} &= \frac{1}{(n-2)(n-3)h^{L+1}} \sum_{k_1 \neq i_1, j_1} \sum_{k_2 \neq i_1, j_1, k_1} K_{vx,h} \left[v_{k_1 k_2} - v_{i_1 j_1}, X_{k_1} - X_{i_1}, X_{k_2} - X_{j_1} \right] \\
\widehat{f}_{x,i_1 j_1} &= \frac{1}{(n-2)(n-3)h^{L}} \sum_{k_1 \neq i_1, j_1} \sum_{k_2 \neq i_1, j_1, k_1} K_{x,h} \left[X_{k_1} - X_{i_1}, X_{k_2} - X_{j_1} \right],
\end{aligned}
$$

where h denotes a bandwith parameter and $L = 2K$. The kernels $K_{vx,h}$ and $K_{x,h}$ are defined as

$$
\begin{aligned}
K_{vx,h} \left[v_{k_1 k_2} - v_{i_1 j_1}, X_{k_1} - X_{i_1}, X_{k_2} - X_{j_1} \right] &= K_{vx} \left[\frac{v_{k_1 k_2} - v_{i_1 j_1}}{h}, \frac{X_{k_1} - X_{i_1}}{h}, \frac{X_{k_2} - X_{j_1}}{h} \right] \\
K_{x,h} \left[X_{k_1} - X_{i_1}, X_{k_2} - X_{j_1} \right] &= K_{x} \left[\frac{X_{k_1} - X_{i_1}}{h}, \frac{X_{k_2} - X_{j_1}}{h} \right].
\end{aligned}
$$

Assumption 4.1.5 below describes the conditions imposed on the kernel functions $K_{vx,h}$ and $K_{x,h}$, and bandwith parameter h.

The estimator defined in equation (9) represents, to the best of my knowledge, the first effort to estimate the vector of parameters θ_0 defined in the network formation model given by equation (3) using a two-step semiparametric estimator that utilizes the existence of a special regressor.

A semiparametric approach is attractive because it does not restrict the distribution of the disturbance term to any specific parametric family. Furthermore, it allows for a flexible statistical dependence between the agent-specific unobserved factors and the observed attributes, i.e., $\{\mathbf{X}_n, \mathbf{A}_n\}$. The estimator defined in equation (9) has as an additional appealing property that it has an analytical form. This characteristic increases its computational tractability compared to the estimator defined as the maximizer of a U-process and introduced in section 3.2. Regarding the non-parametric first-stage estimator, Leung (2015b, Supp. Appendix) and Graham et al. (2019) have studied the properties of kernel estimators for network data. I use their finding to analyze the asymptotic properties of $\widehat{\theta}_n$.

The following technical conditions are needed to prove Theorems 4.1 and 4.2. For simplicity, the theorems are stated and proved assuming that all of the elements of X_i are continuously distributed.

However, the results can be readily extended to include discretely distributed variables by applying the density estimator separately to each discrete cell of data.

Assumption 4.1.1. *For any distinct indices i and j in $\mathcal{N}_n$, the dyad-level covariates (X_i, X_j) and (v_{ij}, X_i, X_j) are absolutely continuous with respect to some Lebesgue measures with Radon-Nikodym densities $f_{x,ij}$ and $f_{vx,ij}$, and supports denoted by $\mathbb{S}_x$ and $\mathbb{S}_{vx}$. Assume that $f_{x,ij}$ and $f_{vx,ij}$ are bounded, $f_{vx,ij}$ is bounded away from zero, and there exists a constant $\overline{M} > L+1$ (recall that $L = 2^K$, with dim$(X_i) = K$) such that $f_{x,ij}$ and $f_{vx,ij}$ are $\overline{M}$-times differentiable with respect to all of its arguments with bounded derivatives. There exist finite constants $C_{w,1}$ and $C_{w,2}$ such that $\sup_{\sigma \in \mathcal{N}_{m_n}} \| \tilde{W}_\sigma \| \leq C_{w,1}$ w.p.1 and $\mathbb{E}\left[\| \tilde{W}_\sigma \|^4\right] < C_{w,2}$.*

Assumption 4.1.1 ensures that the densities $f_{x,ij}$ and $f_{vx,ij}$ are continuous and M-times differentiable. Also, it requires the existence of fourth-order moments for $\tilde{W}_\sigma$, for any $\sigma \in \mathcal{N}_{m_n}$. This assumption has been used in the literature of semiparametric methods, for example in Ahn and Powell (1993); Aradillas-Lopez (2012), and Honoré and Lewbel (2002).

Assumption 4.1.2. *Let τ be a density trimming parameter defined above. Assume that the support $\mathbb{S}_{vx}$ is known, and the the trimming function $I_{\tau,ij}$ is equal to zero if (v_{ij}, X_i, X_j) is within a distance τ of the boundary of $\mathbb{S}_{vx}$, and otherwise, $I_{\tau,ij}$ equals one. Also, assume that $\tau \to 0$ and $\tau n^2 \to 0$ as $n \to \infty$.*

Due to the weighting scheme used in the definition of $\widehat{D}^*_{i_1 j_1}$, boundary effects could arise from the density estimation step when computing $\widehat{\Psi}_{n,\tau}$. Assumptions 4.1.1 and 4.1.2 deal with this technicality by assuming that $f_{vx,i_1 j_1}$ is bounded away from zero and introducing a trimming sequence $I_\tau(v_{i_1 j_1}, X_{i_1}, X_{j_1})$ that sets to zero the terms in $\widehat{\Psi}_{n,\tau}$ with data within a τ distance of the boundary of $\mathbb{S}_{vx}$, (see, e.g., Lewbel 1997, 2000; Honoré and Lewbel 2002, and Khan and Tamer 2010)

Assumption 4.1.1 and Assumption 4.1.2 require that the support $\mathbb{S}_{vx}$ is known. The support $\mathbb{S}_{vx}$ is identified from the distribution of observables, and hence, it can be estimated in an empirical application. As an alternative approach, a fixed trimming function that is not n-dependent could be used instead, (see, e.g., Aradillas-Lopez, Honoré, and Powell 2007 and Aradillas-Lopez 2012).

Assumption 4.1.3. *Let $\overline{M}$ be as defined above. Given any tetrad $\sigma\{i_1, j_1; i_2, j_2\} \in \mathcal{N}_{m_n}$, let*

$$\begin{aligned} \Xi\left(X_{l_1}, X_{l_2}\right) &= E\left[\tilde{W}_\sigma D^*_{l_1 l_2,\tau} \mid X_{l_1}, X_{l_2}\right] \\ \Xi\left(v_{l_1 l_2}, X_{l_1}, X_{l_2}\right) &= E\left[\tilde{W}_\sigma D^*_{l_1 l_2,\tau} \mid v_{l_1 l_2}, X_{l_1}, X_{l_2}\right] \end{aligned}$$

for any dyad $(l_1, l_2) \in \{(i_1, j_1), (i_1, j_2), (i_2, j_1), (i_2, j_2)\}$. The expectations $\Xi\left(x, x\right)$ and $\Xi\left(v, x, x\right)$ exist and are continuous in the components of (v, x, x') for all $(v, x, x') \in \mathbb{S}_{vx}$. Also, $\Xi\left(x, x\right)$ and $\Xi\left(v, x, x\right)$ are $\overline{M}$-times differentiable in the components of (v, x, x') for all $(v, x, x'') \in \overline{\mathbb{S}}_{vx}$, where $\overline{\mathbb{S}}_{vx}$ differs from $\mathbb{S}_{vx}$ by a set of measure zero.

There exist some functions $m_x(x,x)$ and $m_{vx}(v,x,x')$ such that the following local Lipschitz conditions hold for some (x_0,x_0') and (v_0,x_0,x_0') in an open neighborhood of zero and for all $\tau>0$:

$$\begin{aligned}
\|\, f_{vx}(v+v_0,x+x_0,x'+x_0')-f_{vx}(v,x,x')\,\| &\leq m_{vx}(v,x,x')\,\|\,(v_0,x_0,x_0')\,\| \\
\|\, f_x(x+x_0,x'+x_0')-f_x(x,x')\,\| &\leq m_x(x,x')\,\|\,(x_0,x_0')\,\| \\
\|\, \Xi(v+v_0,x+x_0,x'+x_0')-\Xi(v,x,x')\,\| &\leq m_{vx}(v,x,x')\,\|\,(v_0,x_0,x_0')\,\| \\
\|\, \Xi(x+x_0,x'+x_0')-\Xi(x,x')\,\| &\leq m_x(x,x')\,\|\,(x_0,x_0')\,\|\,.
\end{aligned}$$

Assumption 4.1.3 imposes local smoothness conditions that are needed to derive the Hájek projection of a V-statistic. Similar conditions have been used in Ahn and Powell (1993); Aradillas-Lopez (2012), and Honoré and Lewbel (2002).

Assumption 4.1.4. *Given any $\sigma\{i_1,j_1;i_2,j_2\}\in\mathcal{N}_{m_n}$ and $(l_1,l_2)\in\{(i_1,j_1),(i_1,j_2),(i_2,j_1),(i_2,j_2)\}$, let $\chi_{l_1l_2}=\chi(X_{l_1},X_{l_2})=\mathbb{E}\left[\tilde{W}_\sigma\mid X_{l_1},X_{l_2}\right]$.*

The following moments exist

$$\begin{aligned}
&\sup_{(x,x')\in\mathbb{S}_x}\chi(x,x') \\
&\sup_{(v,x,x')\in\mathbb{S}_{v,x},\tau\geq 0}\mathbb{E}\left[\left(\frac{\varphi_{l_1l_2,\tau}}{f_{vx}(v,x,x')}\right)^2\mid v,x,x'\right] \\
&\sup_{(v,x,x')\in\mathbb{S}_{v,x},\tau\geq 0}\mathbb{E}\left[\left(\frac{D^*_{l_1l_2,\tau}}{f_{vx}(v,x,x')}\right)^2\mid v,x,x'\right],
\end{aligned}$$

and the objects

$$\begin{aligned}
&\chi(x,x') \\
&\mathbb{E}\left[\left(\frac{\varphi_{l_1l_2,\tau}}{f_{vx}(v,x,x')}\right)^2\mid v,x,x'\right] \\
&\mathbb{E}\left[\left(\frac{D^*_{l_1l_2,\tau}}{f_{vx}(v,x,x')}\right)^2\mid v,x,x'\right]
\end{aligned}$$

are continuous in the components of $(v,x,x')\in\mathbb{S}_{vx}$. Moreover, there exists a finite constant C_χ, such that

$$E\left[\|\,\chi(x,x')^6\,\|\right]\leq C_\chi$$

for any $(x,x')\in\mathbb{S}_x$.

Assumption 4.1.4 ensures the existence and boundedness of the conditional expectations defined above. These conditions are needed to invoke a uniform law of large numbers for V-statistics. The last part of Assumption 4.1.4 guarantees the existence of sixth-order moments, and it will be used to invoke a conditional central limit theorem.

Assumption 4.1.5. *Let $\overline{M}$ and τ be as defined above. The kernel $K_x(x, x') : \mathbb{R}^L \mapsto \mathbb{R}$ and bandwith h used to define the kernel estimator $\hat{f}_x$ satisfy:*

1. *$K_x(x, x') = 0$ for all (x, x') on the boundary of, and outside of, a convex bounded subset of $\mathbb{R}^L$. This subset has an nonempty interior and has the origin as an interior point.*
2. *$K_x(\cdot, \cdot)$ is symmetric around zero, bounded, differentiable, and bias-reducing of order $2\overline{M}$.*
3. *There exists $\overline{\delta} > 0$ such that $n^{1-\overline{\delta}} h^{L+1} \to \infty$, $nh^{\overline{M}} \to 0$, and $h/\tau \to 0$.*

The kernel function $K_{v,x}(v, x, x')$ has all the same properties, replacing (x, x') with (v, x, x').

Assumption 4.1.5 requires the use of a higher-order kernel. This selection is motivated to control the bias induced by using the inverse of $f_{v|x}(v_{i_1 j_1} \mid X_{i_1}, X_{j_1})$ as a weighting function. This assumption has been used by Honoré and Lewbel (2002) and Leung (2015b). Graham et al. (2019) provides a comprehensive treatment of kernel estimation for undirected network data.

Using the assumptions above, it follows that $\widehat{\theta}_n$ defined in equation (6) is a consistent estimator of θ_0. Theorem 4.1 formally states this result.

Theorem 4.1. *Let Assumptions 3.1.1-3.1.5 and 4.1.1-4.1.5 hold. Then $(\widehat{\theta}_n - \theta_0) \xrightarrow{p} \mathbf{0}$ as $n \to \infty$.*

Proof. See Appendix A. □

The following theorem derives the asymptotic distribution of $\hat{\theta}_n$. A key part to prove this result is to show that

$$\sqrt{n(n-1)}\Upsilon_n^{-1/2} \left\{ \widehat{\Psi}_{n,\tau} - E\left[\tilde{W}_\sigma \widetilde{D}^*_{\sigma,\tau} \mid v_\sigma, X_\sigma, A_\sigma \right] \right\} \Rightarrow \mathcal{N}(0, I),$$

where I denotes the K-dimensional identity matrix, and $\Upsilon_n = n(n-1)Var\left(\widehat{\Psi}_{n,\tau}\right)$, which is defined as

$$\Upsilon_n \quad = \quad \frac{1}{n(n-1)} \sum_{i_1=1}^{n} \sum_{j_1 \neq i_1} \mathbb{E}\left[\left\{ \frac{p_n(\omega_{i_1 j_1})\left[1 - p_n(\omega_{i_1 j_1})\right]}{f_{v|x, i_1 j_1}} \right\} I_{\tau, i_1 j_1} \right] \overline{\chi}_{i_1 j_1} \overline{\chi}'_{i_1 j_1}$$

with $\overline{\chi}_{i_1 j_1} = \left\{ \frac{1}{(n-2)(n-3)} \sum_{i_2 \neq i_1, j_1} \sum_{j_2 \neq i_1, j_1, i_2} \mathbb{E}\left[\tilde{W}_{\sigma\{i_1, i_2; j_1, j_2\}} \mid X_{i_1}, X_{j_1} \right] \right\}$.

The proof of this result follows from showing that

$$\left\{ \widehat{\Psi}_{n,\tau} - \mathbb{E}\left[\tilde{W}_\sigma \widetilde{D}^*_{\sigma,\tau} \mid v_\sigma, X_\sigma, A_\sigma \right] \right\}$$

is asymptotically equivalent to its Hájek Projection onto an arbitrary function of

$$\zeta_{i_1j_1} = (v_{i_1j_1}, X_{i_1}, X_{j_1}, A_{i_1}, A_{j_1}, U_{i_1j_1}).$$

The resulting Hájek Projection is an average of conditionally independent random variables at a dyad level, with conditional mean equal to 0 and a conditional variance that approximates Υ_n in the limit. The result follows from a conditional version of Lyapunov's central limit theorem, (see, e.g., Rao 2009).

The remaining information needed to derive the limiting distribution of the semiparametric estimator $\hat{\theta}_n$, are the convergence rate of Υ_n, which is given by

$$\varrho_n \quad = \quad O\left(\Upsilon_n\right) = O\left(\mathbb{E}\left[\left\{\frac{p_n(\omega_{i_1j_1})\left[1-p_n(\omega_{i_1j_1})\right]}{f_{v|x,i_1j_1}}\right\} I_{\tau,i_1j_1}\right]\right),$$

and the following matrix

$$\Sigma_n \quad = \quad \Gamma_0^{-1} \times \Upsilon_n \times \Gamma_0^{-1}.$$

The next theorem formalizes the limiting distribution of $\widehat{\theta}_n$.

Theorem 4.2. *Suppose Assumptions 3.1.1-3.1.5, 4.1.1-4.1.5, and $n(n-1)\varrho_n^{-1} \to \infty$ hold. It then follows that*

$$\sqrt{n(n-1)}\Sigma_n^{-1/2}\left(\widehat{\theta}_n - \theta_0\right) \quad = \quad \Sigma_n^{-1/2} \times \Gamma_0^{-1} \times \left\{\frac{1}{\sqrt{n(n-1)}}\sum_{i_1=1}^{n}\sum_{j_1=i_1}\xi_{i_1j_1,\tau}\right\} + o_p(1) \qquad (10)$$

with

$$\xi_{i_1j_1,\tau} \quad = \quad \left\{D^*_{i_1j_1} - \mathbb{E}\left[D^*_{i_1j_1} \mid \omega_{i_1j_1}\right]\right\} I_{\tau,i_1j_1}\overline{X}_{i_1j_1},$$

and thus,

$$\sqrt{n(n-1)}\Sigma_n^{-1/2}\left(\widehat{\theta}_n - \theta_0\right) \quad \Rightarrow \quad N\left(0, I\right).$$

Proof. See Appendix A. □

Equation (10) describes the asymptotic linear representation of $\widehat{\theta}_n$ The limiting distribution of $\widehat{\theta}_n$ is derived following a studentized approach as in Andrews and Schafgans (1998), Khan and Tamer (2010), and Jochmans (2018) to control for the possible varying rates of convergence due to sparsity of the network. Notice that if ϱ_n^{-1} converges to a finite constant that is bounded away from zero, $\widehat{\theta}_n - \theta_0$ converges at a parametric rate $\sqrt{n(n-1)}$, with effective sample given by the square root of the number of dyads. Alternatively, if ϱ_n^{-1} decays as n increases, $\widehat{\theta}_n - \theta_0$ has a slower rate of convergence given by $O_p\left(\sqrt{n(n-1)\varrho_n^{-1}}\right)$.

5 Simulations

This section presents simulation evidence for the finite sample performance of the semiparametric estimator introduced in Section 4. I explore the properties of the estimation technique under a wide array of DGP designs that are meant to capture differences in the sample size and in the level of sparsity of the network (Jochmans 2018; Dzemski 2019; Yan et al. 2019).

The undirected network is simulated according the network model in equation (3). I consider a single observed attribute in X_i, which is drawn as $X_i \sim \text{Beta}(2,2) - \frac{1}{2}$. The pair-specific covariate $W_{ij} = g_0(X_i, X_j)$ is constructed to account for complementarities on the observed attributes and is defined as $W_{ij} = X_i X_j$. The agent-specific unobserved factor A_i is generated such that it is correlated with X_i and depends on the sample size n. This last feature offers a useful approach to control the degree of sparsity in the network. In particular, I set

$$A_i \quad = \quad \lambda X_i - (1-\lambda) C_n \times \text{Beta}(0.5, 0.5)$$

where the Beta random variable is independent of X_i and concentrates mass at the boundary of the unit interval. This implies that, conditional on X_i, the individuals cluster at small or high types of unobserved attributes. The parameter $\lambda \in (0,1)$ controls the degree of correlation between the agent-specific heterogeneity and the observed covariate X_i, which is set to $\lambda = \frac{3}{4}$. The constant C_n depends on the size of the network and takes the values $C_n \in \{\log(\log(n)), \log(n)^{1/2}, \log(n)\}$. Under this design, the choice of C_n regulates the degree of sparsity of the network. In fact, for larger values of C_n, fewer links are formed in the network. The special regressor v_{ij} is simulated as $v_{ij} \sim N(0,2)$ for $i < j$, and thus satisfies the support and independence conditions in Assumptions 3.1.3 and 3.1.4. The link-specific disturbance term is generated as $U_{ij} \sim Beta(2,2) - \frac{1}{2}$ for $i < j$.

To simplify the exposition, I focus on the case in which the conditional distribution of v_{ij} is known and consider a fixed trimming design given by $I_{\tau,ij} = 1[\mid v_{ij} \mid < \tau]$ with $\tau = 2std(v_{ij})$. The true DGP is completed by setting $\theta_0 = 1.5$ and consider two different network sizes $n \in \{50, 100\}$.

Table 1 summarizes the results of computing the semiparametric estimator over 500 Monte Carlo replications for all the designs. In particular, I report the mean, median, standard deviation, and mean square error of $\widehat{\theta}_n$ over the total number of simulations. The final column of Table 1 reports the average degree of the network across the total number of simulations. This information will be used to describe the degree of sparsity in the network across the different designs.

The top panel in Table 1 shows the results of estimating θ_0 in a small network with $n = 50$.

Both the mean and the median show that the estimator approximates well the true value of $\theta_0 = 1.5$ independently of the degree of sparsity in the network. Furthermore, these results suggests that the estimator $\widehat{\theta}_n$ presents the smallest dispersion in the dense network design, with $C_n = \log(\log(n))$ and an average degree of 42%. As fewer links are present in the network, the performance of the estimator deteriorates.

In the bottom panel of Table 1, I show the results of estimating θ_0 in a large network with $n = 100$. The evidence in this scenario reinforces the previous findings and suggests that the performance of the estimator $\widehat{\theta}_n$ improves across all the designs. For example, in the dense scenario $C_n = \log(\log(n))$, the standard deviation decreases by an order of less than one half and the mean square error by an order greater than one third. A similar conclusion is obtained from the sparse network case $C_n = \log(n)$ where only 28% of the links are formed. Overall these numerical experiments suggest that the semiparametric estimator $\widehat{\theta}_n$ yields reliable inference for the preference parameter θ_0 in an undirected network formation model.

Table 1: Simulation results for the semiparametric estimator $\widehat{\theta}_n$

$\mid C_n \mid$	mean	median	std	MSE	Degree
		$n = 50$			
$\log(\log(n))$	1.4764	1.4627	0.9158	0.8393	0.4250
$\log(n)^{1/2}$	1.4680	1.4626	1.0672	1.1400	0.3976
$\log(n)$	1.5217	1.6001	1.3832	1.9136	0.3131
		$n = 100$			
$\log(\log(n))$	1.5212	1.5022	0.4809	0.2317	0.4204
$\log(n)^{1/2}$	1.5109	1.5125	0.5687	0.3236	0.3849
$\log(n)$	1.5057	1.4979	0.6916	0.4783	0.2893

[1] Total number of Monte Carlo simulations = 500.

6 Conclusion

This paper has studied a network formation model with unobserved agent-specific heterogeneity. This paper offers two main contributions to the literature on network formation. The first contribution is to propose a new identification strategy that identifies the vector of coefficients θ_0, which accounts for the preferences for homophilic relationships on the observed attributes. The point identification result relies on the existence of a special regressor. This study represents, to the best of my knowledge, the first generalization of a special regressor to analyze a network formation model (Lewbel 1998 and Lewbel 2000).

The second contribution is to introduce a two-step semiparametric estimator for θ_0. The estimator has a closed-form and is computationally tractable even in large networks. I show in Monte Carlo simulations that the estimator performs well in finite samples, as well as in sparse and dense networks.

Two different strands of the literature on network formation have highlighted the importance of accounting for (i) network externalities, and (ii) general forms of unobserved heterogeneity, (see, e.g., Graham 2019b). In future research, I plan to explore the identification power that the special regressor has when considering an augmented model of network formation with network externalities and general forms of unobserved heterogeneity.

References

Ahn, H. and J. L. Powell (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics 58*(1-2), 3–29.

Andrews, D. W. and M. M. Schafgans (1998). Semiparametric estimation of the intercept of a sample selection model. *The Review of Economic Studies 65*(3), 497–517.

Aradillas-Lopez, A. (2010). Semiparametric estimation of a simultaneous game with incomplete information. *Journal of Econometrics 157*(2), 409–431.

Aradillas-Lopez, A. (2012). Pairwise-difference estimation of incomplete information games. *Journal of Econometrics 168*(1), 120–140.

Aradillas-Lopez, A., B. E. Honoré, and J. L. Powell (2007). Pairwise difference estimation with nonparametric control variables. *International Economic Review 48*(4), 1119–1158.

Arellano, M. and B. Honoré (2001). Panel data models: some recent developments. *Handbook of Econometrics 5*, 3229–3296.

Auerbach, E. (2019). Identification and estimation of a partially linear regression model using network data. *arXiv preprint arXiv:1903.09679*.

Badev, A. (2018). Nash equilibria on (un)stable networks.

Boucher, V. and I. Mourifié (2017). My friend far, far away: a random field approach to exponential random graph models. *The econometrics journal 20*(3), S14–S46.

Candelaria, L. E. and T. Ura (2018). Identification and inference of network formation games with misclassified links. *arXiv preprint arXiv:1804.10118*.

Chandrasekhar, A. G. and M. O. Jackson (2014). Tractable and consistent random graph models. *Working Paper*.

Charbonneau, K. B. (2017). Multiple fixed effects in binary response panel data models. *The Econometrics Journal 20*(3), S1–S13.

Chen, S., S. Khan, and X. Tang (2019). Exclusion restrictions in dynamic binary choice panel data models: Comment on "semiparametric binary choice panel data models without strictly exogenous regressors". *Econometrica 87*(5), 1781–1785.

Collomb, G. and W. Härdle (1986). Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations. *Stochastic processes and their applications 23*(1), 77–89.

de Paula, A., S. Richards-Shubik, and E. Tamer (2017). Identifying preferences in networks with bounded degree. forthcoming in. *Econometrica*.

Dzemski, A. (2019). An empirical model of dyadic link formation in a network with unobserved heterogeneity. *Review of Economics and Statistics 101*(5), 763–776.

Gao, W. Y. (2020). Nonparametric identification in index models of link formation. *Journal of Econometrics 215*(2), 399–413.

Gao, W. Y., M. Li, and S. Xu (2020). Logical differencing in dyadic network formation models with nontransferable utilities. *arXiv preprint arXiv:2001.00691*.

Goldsmith-Pinkham, P. and G. W. Imbens (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics 31*(3), 253–264.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica 85*(4), 1033–1063.

Graham, B. S. (2019a). Dyadic regression.

Graham, B. S. (2019b). Network data.

Graham, B. S., F. Niu, and J. L. Powell (2019). Kernel density estimation for undirected dyadic data. *arXiv preprint arXiv:1907.13630*.

Gualdani, C. (2020). An econometric model of network formation with an application to board interlocks between firms.

Han, A. K. (1987). Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator. *Journal of Econometrics 35*(2), 303–316.

Honoré, B. E. and A. Lewbel (2002). Semiparametric binary choice panel data models without strictly exogeneous regressors. *Econometrica 70*(5), 2053–2063.

Jackson, M. O. and A. Wolinsky (1996). A strategic model of social and economic networks. *Journal of economic theory 71*(1), 44–74.

Jochmans, K. (2017). Two-way models for gravity. *Review of Economics and Statistics 99*(3), 478–485.

Jochmans, K. (2018). Semiparametric analysis of network formation. *Journal of Business & Economic Statistics 36*(4), 705–713.

Khan, S. and E. Tamer (2010). Irregular identification, support conditions, and inverse weight estimation. *Econometrica 78*(6), 2021–2042.

Lee, A. J. (2019). *U-statistics: Theory and Practice*. Routledge.

Leung, M. (2015a). A random-field approach to inference in large models of network formation. *Available at SSRN*.

Leung, M. (2015b). Two-step estimation of network-formation models with incomplete information. *Journal of Econometrics 188*(1), 182–195.

Lewbel, A. (1997). Semiparametric estimation of location and other discrete choice moments. *Econometric Theory 13*(01), 32–51.

Lewbel, A. (1998). Semiparametric latent variable model estimation with endogenous or mismeasured regressors. *Econometrica*, 105–121.

Lewbel, A. (2000). Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics 97*(1), 145–177.

Lewbel, A. (2012). *An overview of the special regressor method.* Boston College, Department of Economics.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics 3*(3), 205–228.

Manski, C. F. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of Econometrics 27*(3), 313–333.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 415–444.

Mele, A. (2017). A structural model of dense network formation. *Econometrica 85*(3), 825–850.

Menzel, K. (2015). Strategic network formation with many agents.

Miyauchi, Y. (2016). Structural estimation of a pairwise stable network with nonnegative externality. *Journal of Econometrics, Forthcoming*.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics 4*, 2111–2245.

Powell, J. L. (1994). Estimation of semiparametric models. *Handbook of econometrics 4*, 2443–2521.

Powell, J. L., J. H. Stock, and T. M. Stoker (1989). Semiparametric estimation of index coefficients. *Econometrica: Journal of the Econometric Society*, 1403–1430.

Rao, B. P. (2009). Conditional independence, conditional mixing and conditional association. *Annals of the Institute of Statistical Mathematics 61*(2), 441–460.

Ridder, G. and S. Sheng (2020). Estimation of large network formation games. *arXiv preprint arXiv:2001.03838*.

Serfling, R. J. (2009). *Approximation theorems of mathematical statistics*, Volume 162. John Wiley & Sons.

Sheng, S. (2018). A structural econometric analysis of network formation games through subnetworks. *forthcoming in Econometrica, mimeo UCLA*.

Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *The Annals of Statistics*, 177–184.

Toth, P. (2017). Semiparametric estimation in networks with homophily and degree heterogeneity. Technical report, Working paper, University of Nevada.

Yan, T., B. Jiang, S. E. Fienberg, and C. Leng (2019). Statistical inference in a directed network model with covariates. *Journal of the American Statistical Association 114*(526), 857–868.

Zeleneev, A. (2020). Identification and estimation of network models with nonparametric unobserved heterogeneity.

A Appendix

A.1 Proof of Theorem 3.1

Proof. Let $e_{ij} = A_i + A_j - U_{ij}$ and $s(w,e) = -w'\theta_0 - e$. Consider

$$\begin{aligned}
E[D^*_{ij} \mid X_i, X_j] &= E\left[E\left[D^*_{ij} \mid v_{ij}, X_i, X_j\right] \mid X_i, X_j\right] \\
&= \int_{\underline{s}_v}^{\overline{s}_v} \frac{E\left[D_{ij} - \mathbf{1}\left[v_{ij} > 0\right] \mid v_{ij}, X_i, X_j\right]}{f_{v|x}(v_{ij} \mid X_i, X_j)} f_{v|x}(v_{ij} \mid X_i, X_j)\, dv_{ij} \\
&= \int_{\underline{s}_v}^{\overline{s}_v} E\left[\mathbf{1}\left[v_{ij} \geq s(W_{ij}, e_{ij})\right] - \mathbf{1}\left[v_{ij} > 0\right] \mid v_{ij}, X_i, X_j\right]\, dv_{ij} \\
&= \int_{\underline{s}_v}^{\overline{s}_v} \int_{\mathbb{S}_e(X_i,X_j)} \left\{\mathbf{1}\left[v_{ij} \geq s(W_{ij}, e_{ij})\right] - \mathbf{1}\left[v_{ij} > 0\right]\right\} dF_{e|x}(e_{ij} \mid v_{ij}, X_i, X_j)\, dv_{ij} \\
&= \int_{\mathbb{S}_e(X_i,X_j)} \int_{\underline{s}_v}^{\overline{s}_v} \left\{\mathbf{1}\left[v_{ij} \geq s(W_{ij}, e_{ij})\right] - \mathbf{1}\left[v_{ij} > 0\right]\right\} dv_{ij}\, dF_{e|x}(e_{ij} \mid X_i, X_j) \\
&= \int_{\mathbb{S}_e(X_i,X_j)} -s(W_{ij}, e_{ij}) dF_{e|x}(e_{ij} \mid X_i, X_j) \\
&= \int_{\mathbb{S}_e(X_i,X_j)} \left(W'_{ij}\theta_0 + e_{ij}\right) dF_{e|x}(e_{ij} \mid X_i, X_j) \\
&= W'_{ij}\theta_0 + E\left[e_{ij} \mid X_i, X_j\right].
\end{aligned}$$

The third to last equality follows from the next result

$$\begin{aligned}
\int_{\underline{s}_v}^{\overline{s}_v} \left\{\mathbf{1}\left[v_{ij} \geq s(W_{ij}, e_{ij})\right] - \mathbf{1}\left[v_{ij} > 0\right]\right\} dv_{ij} &= \int_{s(W_{ij},e_{ij})}^{\overline{s}_v} 1 dv_{ij} - \overline{s}_v \\
&= -s(W_{ij}, e_{ij}).
\end{aligned}$$

□

A.2 Proof of Corollary 3.1

Proof. Theorem 3.1 concludes that

$$E\left[D^*_{ik} \mid X_i, X_k\right] = W'_{ik}\theta_0 + E\left[A_i + A_k \mid X_i, X_k\right].$$

Observe that D^*_{ik} is a function of $(Z_i, Z_k, A_i, A_k, U_{ik})$. It follows from the the random sampling of nodes, Assumption 3.1.1, and the conditionally independent formation of links, Assumption 3.1.2, that the following condition holds for any tetrad $\sigma\{i,j,k,l\} \in \mathcal{N}_{m_n}$

$$\begin{aligned}
E[D^*_{ik} \mid X_i, X_k] &= E[D^*_{ik} \mid X_{\sigma(\{i,j,k,l\})}] \\
E\left[A_i + A_k \mid X_i, X_k\right] &= E\left[A_i + A_k \mid X_{\sigma(\{i,j,k,l\})}\right],
\end{aligned}$$

since the joint distribution of $(v_i, v_k, A_i, A_k, U_{ik})$ is conditionally independent of (X_j, X_l), given (X_i, X_k), i.e.,

$$
\begin{aligned}
Pr(v_i, v_k, A_i, A_k, U_{ik} \mid X_i, X_k) &= Pr(U_{ik} \mid X_i, X_k, v_i, v_k, A_i, A_k) Pr(v_i, v_k, A_i, A_k \mid X_i, X_k) \\
&= Pr(U_{ik} \mid X_{\sigma(\{i,j,k,l\})}, v_i, v_k, A_i, A_k) Pr(v_i, v_k, A_i, A_k \mid X_{\sigma(\{i,j,k,l\})}) \\
&= Pr(v_i, v_k, A_i, A_k, U_{ik} \mid X_{\sigma(\{i,j,k,l\})}),
\end{aligned}
$$

where the second equality follows from Assumptions 3.1.1 and 3.1.2. Thus, the results above yield

$$
\begin{aligned}
E[D^*_{ik} - D^*_{il} \mid X_{\sigma(\{i,j,k,l\})}] &= (W_{ik} - W_{il})'\theta_0 + E\left[A_k - A_l \mid X_{\sigma(\{i,j,k,l\})}\right] \\
E[D^*_{jk} - D^*_{jl} \mid X_{\sigma(\{i,j,k,l\})}] &= (W_{jk} - W_{jl})'\theta_0 + E\left[A_k - A_l \mid X_{\sigma(\{i,j,k,l\})}\right],
\end{aligned}
$$

for any tetrad $\sigma\{i, j, k, l\}$, which in turn implies

$$
E[\tilde{D}^*_\sigma \mid X_\sigma] = \tilde{W}'_\sigma \theta_0. \tag{11}
$$

The result follows from Assumption 3.1.5. The proof is complete. □

Proof of Theorem 3.2

Proof. First, notice that for any $(X_\sigma, \bar{\boldsymbol{v}}_\sigma) \in Q_\theta$

$$
\text{sign}\{\tilde{v}_\sigma\} = \text{sign}\left\{\tilde{v}_\sigma + \left(\Delta_\sigma W'_i\theta_0 + \Delta_\sigma A\right) - \left(\Delta_\sigma W'_j\theta_0 + \Delta_\sigma A\right)\right\}
$$

since $\mid \tilde{v}_\sigma \mid \geq \overline{s}_\varepsilon - \underline{s}_\varepsilon$ with probability 1.

Consider a $\theta \neq \theta_0$ with $P\left[\bar{\boldsymbol{Z}}_\sigma \in Q_\theta\right] > 0$. Without loss of generality, consider some $(X_\sigma, \bar{\boldsymbol{v}}_\sigma) \in Q_\theta$, with $\tilde{W}'_\sigma\theta \leq -\tilde{v}_\sigma < \tilde{W}'_\sigma\theta_0$. From the previous observation, it follows that $\tilde{v}_\sigma + \tilde{W}'_\sigma\theta_0 + \Delta_\sigma A - \Delta_\sigma A > 0$ and $\Delta_\sigma v_i > \overline{s}_\varepsilon$, $\Delta_\sigma v_j < \underline{s}_\varepsilon$ with probability 1.

Given $(X_\sigma, \bar{\boldsymbol{v}}_\sigma) \in Q_\theta$, it follows that $\tilde{v}_\sigma + \tilde{W}'_\sigma\theta_0 + \Delta_\sigma A - \Delta_\sigma A > 0$ and $\Delta_\sigma v_i > \overline{s}_\varepsilon$, $\Delta_\sigma v_j < \underline{s}_\varepsilon$ hold if and only if

$$
\begin{aligned}
\Delta_\sigma v_i &> -\left(\Delta_\sigma W'_i\theta_0 + \Delta_\sigma A\right) \\
\Delta_\sigma v_j &\leq -\left(\Delta_\sigma W'_j\theta_0 + \Delta_\sigma A\right)
\end{aligned} \tag{12}
$$

with probability 1. The inequalities in (12) are sufficient conditions for

$$
\begin{aligned}
& P_{\theta_0}\left[\tilde{D}_\sigma = 2 \mid X_\sigma, A_\sigma, \bar{\boldsymbol{v}}_\sigma \in \mathcal{V}(X_\sigma), \tilde{D}_\sigma \in \{-2, 2\}\right] \\
> \; & P_{\theta_0}\left[\tilde{D}_\sigma = -2 \mid X_\sigma, A_\sigma, \bar{\boldsymbol{v}}_\sigma \in \mathcal{V}(X_\sigma), \tilde{D}_\sigma \in \{-2, 2\}\right],
\end{aligned}
$$

or equivalently, for

$$
\mathbb{E}_{\theta_0}\left[\tilde{D}_\sigma \mid X_\sigma, A_\sigma, \bar{\boldsymbol{v}}_\sigma \in \mathcal{V}(X_\sigma), \tilde{D}_\sigma \in \{-2, 2\}\right] > 0.
$$

Notice that for a $(X_\sigma, \bar{\boldsymbol{v}}_\sigma) \in Q_\theta$, $\mathbb{E}_{\theta_0}\left[\tilde{D}_\sigma \mid X_\sigma, A_\sigma, \bar{\boldsymbol{v}}_\sigma \in \mathcal{V}(X_\sigma), \tilde{D}_\sigma \in \{-2,2\}\right] > 0$ is also sufficient to conclude that $\tilde{v}_\sigma + \tilde{W}_\sigma'\theta_0 + \Delta_\sigma A - \Delta_\sigma A > 0$ with probability 1. Otherwise, if $\tilde{v}_\sigma + \tilde{W}_\sigma'\theta_0 + \Delta_\sigma A - \Delta_\sigma A \leq 0$ with $\bar{\boldsymbol{v}}_\sigma \in \mathcal{V}(X_\sigma)$, it would be the case that $\tilde{v}_\sigma < 0$, and thus

$$\begin{aligned} \Delta_\sigma v_i &\leq -\left(\Delta_\sigma W_i'\theta_0 + \Delta_\sigma A\right) \\ \Delta_\sigma v_j &> -\left(\Delta_\sigma W_j'\theta_0 + \Delta_\sigma A\right) \end{aligned}$$

with probability 1, which contradicts

$$\mathbb{E}_{\theta_0}\left[\tilde{D}_\sigma \mid X_\sigma, A_\sigma, \bar{\boldsymbol{v}}_\sigma \in \mathcal{V}(X_\sigma), \tilde{D}_\sigma \in \{-2,2\}\right] > 0.$$

Hence,

$$\text{sign}\left\{\mathbb{E}_{\theta_0}\left[\tilde{D}_\sigma \mid X_\sigma, A_\sigma, \bar{\boldsymbol{v}}_\sigma \in \mathcal{V}(X_\sigma), \tilde{D}_\sigma \in \{-2,2\}\right]\right\} = \text{sign}\left\{\tilde{v}_\sigma + \tilde{W}_\sigma'\theta_0\right\}$$

for any $(X_\sigma, A_\sigma, \bar{\boldsymbol{v}}_\sigma \in \mathcal{V}(X_\sigma))$.

The previous result implies that for any $(X_\sigma, \bar{\boldsymbol{v}}_\sigma) \in Q_\theta$ with $P\left[\bar{\boldsymbol{Z}}_\sigma \in Q_\theta\right] > 0$, it will holds that $\tilde{W}_\sigma'\theta \leq -\tilde{v}_\sigma < \tilde{W}_\sigma'\theta_0$ if and only if

$$\text{sign}\left\{\mathbb{E}_{\theta_0}\left[\tilde{D}_\sigma \mid X_\sigma, \bar{\boldsymbol{v}}_\sigma \in \mathcal{V}, \tilde{D}_\sigma \in \{-2,2\}\right]\right\} > \text{sign}\left\{\mathbb{E}_{\theta}\left[\tilde{D}_\sigma \mid X_\sigma, \bar{\boldsymbol{v}}_\sigma \in \mathcal{V}, \tilde{D}_\sigma \in \{-2,2\}\right]\right\}.$$

This result implies that $\bar{\boldsymbol{z}}_\sigma \in \xi_\theta(X_\sigma)$, and $P\left[\bar{\boldsymbol{Z}}_\sigma \in \xi_\theta\right] > 0$. Therefore, θ_0 is identified relative to θ. □

Proof of Corollary 3.3

Proof. Consider any $\theta \neq \theta_0$. It follows from Assumption 3.2.4 that $P\left[\tilde{W}_\sigma'(\theta - \theta_0) \neq 0\right] > 0$ for any tetrad $\sigma \in \mathcal{N}_{m_n}$. Suppose without loss of generality that $P\left[\tilde{W}_\sigma'\theta < \tilde{W}_\sigma'\theta_0\right] > 0$. Under Assumptions 3.1.1 and 3.2.3, for any X_σ, with $\tilde{W}_\sigma'\theta < \tilde{W}_\sigma'\theta_0$, there exists an interval of $\tilde{v}_\sigma = \Delta_\sigma v_i - \Delta_\sigma v_j$ with $\tilde{W}_\sigma'\theta \leq -\tilde{v}_\sigma < \tilde{W}_\sigma'\theta_0$. This implies that $P\left[\bar{\boldsymbol{Z}}_\sigma \in Q_\theta\right] > 0$, and thus θ_0 is point identified relative to all $\theta \neq \theta_0$. □

A.3 Proof of Theorem 4.1

Proof. Consider $\widehat{\theta}_n = \widehat{\Gamma}_n^{-1} \times \widehat{\Psi}_{n,\tau}$, with

$$\begin{aligned} \widehat{\Gamma}_n &= \frac{1}{m_n}\sum_{\sigma \in \mathcal{N}_{m_n}}\left[\tilde{W}_\sigma \tilde{W}_\sigma'\right] \\ \widehat{\Psi}_{n,\tau} &= \frac{1}{m_n}\sum_{\sigma \in \mathcal{N}_{m_n}}\left[\tilde{W}_\sigma \widehat{D}^*_{\sigma,\tau}\right] \end{aligned}$$

First, I will show that $\widehat{\Gamma}_n \xrightarrow{p} \Gamma_0$ and $\widehat{\Psi}_{n,\tau} \xrightarrow{p} \Psi_0$; the result will follow Assumption 3.1.5, the continuous mapping theorem and Slutsky.

Part 1. Notice that $\widehat{\Gamma}_n - \Gamma_0$ is a mean zero fourth-order V-statistic, without common indices

$$\widehat{\Gamma}_n - \Gamma_0 \quad = \quad \frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} \left\{ \left[\tilde{W}_\sigma \tilde{W}_\sigma' \right] - E\left[\tilde{W}_\sigma \tilde{W}_\sigma' \right] \right\}.$$

Lemma B.1 implies that $\hat{\Gamma}_n - \Gamma_0$ can be approximated by a mean zero U-statistic of order 4 at a rate $\sqrt{n}$. Assumption 3.1.5 ensures that Γ_0 is finite. It follows from Assumption 3.1.1 that a Strong Law of Large Numbers for U-statisitcs holds, and hence, $\hat{\Gamma}_n - \Gamma_0 = o_p(1)$, (see Serfling 2009, Theorem A, p. 190).

Part 2. For a fixed tetrad $\sigma = \sigma\{i_1, i_2; j_1, j_2\} \in \mathcal{N}_{m_n}$, let

$$\widehat{\eta}_{[l_1 l_2],\tau} \quad = \quad \frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} \tilde{W}_\sigma \varphi_{l_1 l_2,\tau} \left(\frac{\hat{f}_{x,l_1 l_2}}{\hat{f}_{vx,l_1 l_2}} \right),$$

for $(l_1, l_2) \in \{(i_1, j_1), (i_1, j_2), (i_2, j_1), (i_2, j_2)\}$. Next, observe that $\widehat{\Psi}_{n,\tau}$ can be written as

$$\widehat{\Psi}_{n,\tau} \quad = \quad \left(\widehat{\eta}_{[i_1 j_1],\tau} - \widehat{\eta}_{[i_1 j_2],\tau} \right) - \left(\widehat{\eta}_{[i_2 j_1],\tau} - \widehat{\eta}_{[i_2 j_2],\tau} \right).$$

Consistent estimation of Ψ_0 will follow from repeated applications of Lemma B.2. It follows from Lemma B.3 that $\widehat{\eta}_{[l_1 l_2],\tau}$ can be written as

$$\widehat{\eta}_{[l_1 l_2],\tau} = \frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} \tilde{W}_\sigma \varphi_{l_1 l_2,\tau} \left\{ \frac{f_{x,l_1 l_2}}{f_{vx,l_1 l_2}} + \frac{\widehat{f}_{x,l_1 l_2} - f_{x,l_1 l_2}}{f_{vx,l_1 l_2}} - \frac{f_{x,l_1 l_2}}{f_{vx,l_1 l_2}} \times \frac{\widehat{f}_{vx,l_1 l_2} - f_{vx,l_1 l_2}}{f_{vx,l_1 l_2}} \right\} + o_p(1).$$

Then, Lemma B.2 yields

$$\begin{aligned} \frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} \tilde{W}_\sigma \varphi_{l_1 l_2,\tau} \left\{ \frac{\hat{f}_{x,l_1 l_2}}{f_{vx,l_1 l_2}} \right\} \quad &= \quad E\left[\tilde{W}_\sigma \varphi_{l_1 l_2,\tau} \frac{f_{x,l_1 l_2}}{f_{vx,l_1 l_2}} \right] + o_p(1) \\ \frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} \tilde{W}_\sigma \varphi_{l_1 l_2,\tau} \left\{ \frac{f_{x,l_1 l_2}}{f_{vx,l_1 l_2}} \left(\frac{\hat{f}_{vx,l_1 l_2}}{f_{vx,l_1 l_2}} \right) \right\} \quad &= \quad E\left[\tilde{W}_\sigma \varphi_{l_1 l_2,\tau} \frac{f_{x,l_1 l_2}}{f_{vx,l_1 l_2}} \right] + o_p(1). \end{aligned}$$

It follows from the previous results and the definition of $D^*_{l_1 l_2,\tau}$ that

$$\widehat{\eta}_{[i_1 j_1],\tau} \quad = \quad \frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} \left\{ \tilde{W}_\sigma D^*_{l_1 l_2,\tau} \right\} + E\left[\tilde{W}_\sigma D^*_{l_1 l_2,\tau} \right] - E\left[\tilde{W}_\sigma D^*_{l_1 l_2,\tau} \right] + o_p(1),$$

which is a V-statistic of order 4. It follows from Lemma B.1 that it can be approximated by a U-statisitcs of order 4. Assumptions 4.1.1 and 4.1.2, and equation (6) ensure that $E\left[\tilde{W}_\sigma D^*_{l_1 l_2,\tau} \right]$ is finite. It follows then from Assumptions 3.1.1 that a Strong Law of Large Numbers for U-statistics holds, and hence,

$$\frac{1}{m_n}\sum_{\sigma\in\mathcal{N}_{m_n}}\left\{\tilde{W}_\sigma D^*_{l_1l_2,\tau}-E\left[\tilde{W}_\sigma D^*_{l_1l_2,\tau}\right]\right\} \;=\; o_p(1).$$

Consider next

$$\frac{1}{m_n}\sum_{\sigma\in\mathcal{N}_{m_n}}\tilde{W}_\sigma\left\{D^*_{l_1l_2}-D^*_{l_1l_2,\tau}\right\} \;=\; \frac{1}{n(n-1)}\sum_{l_1}^{n}\sum_{l_2\neq l_1}D^*_{l_1l_2}\left\{(1-I_{\tau,l_1l_2})\right\}\tilde{W}_{l_1l_2}(\sigma)$$

where the equality follows from the definition of $D^*_{l_1l_2,\tau}$ and

$$\tilde{W}_{l_1l_2}(\sigma) \;=\; \frac{1}{(n-2)(n-3)}\sum_{s_1\neq l_1,l_2}\sum_{s_2\neq l_1,l_2,s_1}\tilde{W}_{\sigma\{l_1,s_1;l_2,s_2\}}.$$

It follows from using a Cauchy-Schwarz inequality, that the expectation

$$E\left[\left(\frac{1}{n(n-1)}\sum_{l_1}^{n}\sum_{l_2\neq l_1}D^*_{l_1l_2}\left\{(1-I_{\tau,l_1l_2})\right\}\tilde{W}_{l_1l_2}(\sigma)\right)^2\right]$$

is bounded by

$$\begin{aligned}\frac{1}{n(n-1)}\sum_{l_1}^{n}\sum_{l_2\neq l_1}E\left[\left(D^*_{l_1l_2}\left\{(1-I_{\tau,l_1l_2})\right\}\tilde{W}_{l_1l_2}(\sigma)\right)^2\right] \;&=\; O\left(E\left[\tilde{W}_{l_1l_2}(\sigma)^2\left(D^*_{l_1l_2}\right)^2(1-I_{\tau,l_1l_2})^2\right]\right)\\ &\leq\; \sup_\sigma\left(\tilde{W}^2_\sigma\right)\sup_{l_1l_2}\left(D^*_{l_1l_2}\right)^2 O\left(E\left[(1-I_{\tau,l_1l_2})^2\right]\right).\end{aligned}$$

where the inequality follows from Assumption 4.1.1. Assumption 4.1.2 yields

$$E\left[(1-I_{\tau,l_1l_2})^2\right] \;=\; P\left[I_{\tau,l_1l_2}=0\right]=o(\tau).$$

Using the results above to conclude that

$$E\left[\left(\frac{1}{n(n-1)}\sum_{l_1}^{n}\sum_{l_2\neq l_1}D^*_{l_1l_2}\left\{(1-I_{\tau,l_1l_2})\right\}\tilde{W}_{l_1l_2}(\sigma)\right)^2\right] \;\leq\; o(\tau),$$

and hence,

$$\frac{1}{m_n}\sum_{\sigma\in\mathcal{N}_{m_n}}\left\{\tilde{W}_\sigma D^*_{l_1l_2,\tau}-E\left[\tilde{W}_\sigma D^*_{l_1l_2}\right]\right\} \;=\; o(1).$$

Using similar steps for $\widehat{\eta}_{[i_1j_2],\tau}$, $\widehat{\eta}_{[i_2j_1],\tau}$, and $\widehat{\eta}_{[i_2j_2],\tau}$, yield

$$\hat{\Psi}_{n,\tau}-E\left[\tilde{W}_\sigma\tilde{D}^*_\sigma\right] \;=\; o_p(1).$$

The result follows from Assumption 3.1.5, the Continuous Mapping Theorem and Slutsky Theorem. □

A.4 Proof of Theorem 4.2

Proof. **Part 1: Hájek Projection**

Under Assumptions 3.1.1-3.1.5, 4.1.1-4.1.5, it follows from the proof of Theorem 4.1 that $\hat{\Gamma}_n \xrightarrow{p} \Gamma_0$, and from Lemma B.3 that $\widehat{\eta}_{[l_1l_2],\tau}$ can be written as

$$\widehat{\eta}_{[l_1l_2],\tau} = \frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} \tilde{W}_\sigma \varphi_{l_1l_2,\tau} \left\{ \frac{f_{x,l_1l_2}}{f_{vx,l_1l_2}} + \frac{\widehat{f}_{x,l_1l_2} - f_{x,l_1l_2}}{f_{vx,l_1l_2}} - \frac{f_{x,l_1l_2}}{f_{vx,l_1l_2}} \times \frac{\widehat{f}_{vx,l_1l_2} - f_{vx,l_1l_2}}{f_{vx,l_1l_2}} \right\} + o_p(1)$$

for $(l_1, l_2) \in \{(i_1, j_1), (i_1, j_2), (i_2, j_1), (i_2, j_2)\}$.

Hence, $\widehat{\Psi}_{n,\tau} = \left(\widehat{\eta}_{[i_1j_1],\tau} - \widehat{\eta}_{[i_1j_2],\tau}\right) - \left(\widehat{\eta}_{[i_2j_1],\tau} - \widehat{\eta}_{[i_2j_2],\tau}\right)$, which can be expressed as $\widehat{\Psi}_{n,\tau} = S_{1,n\tau} + S_{2,n\tau} - S_{3,n\tau} + o_p(1)$ using the expression above, with

$$\begin{aligned}
S_{1,n\tau} &= \frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} \tilde{W}_\sigma \left\{ \left(D^*_{i_1j_1,\tau} - D^*_{i_1j_2,\tau}\right) - \left(D^*_{i_2j_1,\tau} - D^*_{i_2j_2,\tau}\right) \right\} \\
S_{2,n\tau} &= \frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} \tilde{W}_\sigma \left\{ \left(\frac{\varphi_{i_1j_1,\tau}\widehat{f}_{x,i_1j_1}}{f_{vx,i_1j_1}} - \frac{\varphi_{i_1j_2,\tau}\widehat{f}_{x,i_1j_2}}{f_{vx,i_1j_2}} \right) - \left(\frac{\varphi_{i_2j_1,\tau}\widehat{f}_{x,i_2j_1}}{f_{vx,i_2j_1}} - \frac{\varphi_{i_2j_2,\tau}\widehat{f}_{x,i_2j_2}}{f_{vx,i_2j_2}} \right) \right\} \\
S_{3,n\tau} &= \frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} \tilde{W}_\sigma \left\{ \left(\frac{D^*_{i_1j_1,\tau}\widehat{f}_{vx,i_1j_1}}{f_{vx,i_1j_1}} - \frac{D^*_{i_1j_2,\tau}\widehat{f}_{vx,i_1j_2}}{f_{vx,i_1j_2}} \right) - \left(\frac{D^*_{i_2j_1,\tau}\widehat{f}_{vx,i_2j_1}}{f_{vx,i_2j_1}} - \frac{D^*_{i_2j_2,\tau}\widehat{f}_{vx,i_2j_2}}{f_{vx,i_2j_2}} \right) \right\}.
\end{aligned}$$

Consider

$$\left(\widehat{\Psi}_{n,\tau} - E\left[\tilde{W}_\sigma \widetilde{D}^*_{\sigma,\tau} \mid \Omega_n\right]\right) = \left\{S_{1,n\tau} - E\left[\tilde{W}_\sigma \widetilde{D}^*_{\sigma,\tau} \mid \Omega_n\right]\right\} + S_{2,n\tau} - S_{3,n\tau} + o_p(1),$$

it follows from Lemmas B.4, B.5, and B.6 that the Hájek projection of

$$\left(\widehat{\Psi}_{n,\tau} - E\left[\tilde{W}_\sigma \widetilde{D}^*_{\sigma,\tau} \mid \Omega_n\right]\right)$$

into an arbitrary function of $\zeta_{i_1j_1} = (X_{i_1}, X_{j_1}, A_{i_1}, A_{j_1}, v_{i_1j_1}, U_{i_1j_1})$ is given by

$$\left(\widehat{\Psi}_{n,\tau} - E\left[\tilde{W}_\sigma \tilde{D}^*_{\sigma\tau} \mid \Omega_n\right]\right) = V^*_n + o_p\left(\sqrt{\frac{\varrho_n}{n(n-1)}}\right)$$

where

$$
\begin{aligned}
V_n^* &= \frac{1}{n(n-1)} \sum_{i_1=1}^{n} \sum_{j_1 \neq i_1} \xi_{i_1 j_1,\tau} \\
\xi_{i_1 j_1,\tau} &= \left\{ D_{i_1 j_1}^* - E\left[D_{i_1 j_1}^* \mid \omega_{i_1 j_1} \right] \right\} I_{\tau,i_1 j_1} \overline{\chi}_{i_1 j_1} \\
\overline{\chi}_{i_1 j_1} &= \left\{ \frac{1}{(n-2)(n-3)} \sum_{i_2 \neq i_1, j_1} \sum_{j_2 \neq i_1, j_1, i_2} E\left[\tilde{W}_{\sigma\{i_1,i_2;j_1,j_2\}} \mid X_{i_1}, X_{j_1} \right] \right\}
\end{aligned}
$$

and

$$
\begin{aligned}
\Upsilon_{n,\tau} &= n(n-1) Var\left(V_n^*\right) = \frac{1}{n(n-1)} \left\{ \sum_{i_1=1}^{n} \sum_{j_1 \neq i_1} \Lambda_{i_1,j_1}^* \right\} \\
\Lambda_{i_1,j_1}^* &= E\left[\left\{ E\left[D_{i_1 j_1}^* D_{i_1 j_1}^* \mid \omega_{i_1 j_1} \right] - E\left[D_{i_1 j_1}^* \mid \omega_{i_1 j_1} \right]^2 \right\} I_{\tau,i_1 j_1}^2 \overline{\chi}_{i_1 j_1} \overline{\chi}_{i_1 j_1}' \right] \\
\varrho_{n,\tau} &= O\left(\Upsilon_{n,\tau}\right) = O\left(E\left[\left\{ \frac{p_n(\omega_{i_1 j_1})\left[1 - p_n(\omega_{i_1 j_1})\right]}{f_{v|x,i_1 j_1}} \right\} I_{\tau,i_1 j_1} \right] \right).
\end{aligned}
$$

Part 2: Bias Reduction

Consider next,

$$
n(n-1)\varrho_n^{-1} E\left[\left(\frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} \tilde{W}_\sigma \left\{ \tilde{D}_{\sigma\tau}^* - \tilde{D}_\sigma^* \right\} \right)^2 \mid \Omega_n \right].
$$

It follows from a Cauchy-Schwarz inequality that the term above is bounded by

$$
n(n-1)\varrho_n^{-1} \frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} E\left[\left(\tilde{D}_{\sigma\tau}^* - \tilde{D}_\sigma^* \right)^2 \mid \Omega_n \right] \tilde{W}_\sigma \tilde{W}_\sigma'
$$

which is equal to

$$
O\left(n(n-1)\varrho_n^{-1} \left\{ E\left[D_{i_1 j_1}^* D_{i_1 j_1}^* \mid \omega_{i_1 j_1} \right] - E\left[D_{i_1 j_1}^* \mid \omega_{i_1 j_1} \right] E\left[D_{i_1 j_1}^* \mid \omega_{i_1 j_1} \right]' \right\} \left(I_{\tau,i_1 j_1} - 1 \right)^2 \tilde{W}_\sigma \tilde{W}_\sigma' \right).
$$

Assumptions 4.1.1 and 4.1.2 yield

$$
\begin{aligned}
& \sup_\sigma \left(\tilde{W}_\sigma \right) \sup_\sigma \left(\tilde{W}_\sigma \right)' O\left(n(n-1)\varrho_n^{-1} \left\{ \frac{p_n\left(\omega_{i_1 j-1}\right)\left[1 - p_n\left(\omega_{i_1 j-1}\right)\right]}{f_{v|x,i_1 j_1}} \right\} \left(I_{\tau,i_1 j_1} - 1 \right)^2 \right) \\
= \;& O\left(n(n-1)\tau \right) = 0
\end{aligned}
$$

since $(I_{\tau,i_1 j_1} - 1)$ as $\tau \to 0$ and $n \to \infty$.

Therefore,

$$n(n-1)\varrho_{1n}^{-1}E\left[\left(\frac{1}{m_n}\sum_{\sigma\in\mathcal{N}_{m_n}}\tilde{W}_\sigma\left\{\tilde{D}^*_{\sigma\tau}-\tilde{D}^*_\sigma\right\}\right)^2\mid\Omega_n\right] \quad = \quad o(1),$$

and so

$$n(n-1)\varrho_{1n}^{-1}\left(E\left[\tilde{W}_\sigma\tilde{D}^*_{\sigma\tau}\mid\Omega_\sigma\right]-E\left[\tilde{W}_\sigma\tilde{D}^*_\sigma\mid\Omega_\sigma\right]\right) \quad = \quad o(1).$$

Part 3: Limit Distribution of Projection

Given Assumptions 3.1.2, the Hájek projection V_n^* is an average of $\{\xi_{i_1j_1,\tau}\}$, which are conditionally independent given $\Omega_n=(\mathbf{v}_n,\mathbf{X}_n,\mathbf{A}_n)$, with conditional mean

$$E\left[\xi_{i_1j_1,\tau}\mid\Omega_n\right] \quad = \quad 0$$

and conditional variance

$$\begin{aligned}\Upsilon\left(\Omega_n\right) \quad &= \quad n(n-1)Var\left(\frac{1}{n(n-1)}\sum_{i_1=1}^n\sum_{j_1\neq i_1}\xi_{i_1j_1}\mid\Omega_n\right)\\ &= \quad \frac{1}{n(n-1)}\sum_{i_1=1}^n\sum_{j_1\neq i_1}\left\{E\left[D^*_{i_1j_1}D^*_{i_1j_1}\mid\omega_{i_1j_1}\right]-E\left[D^*_{i_1j_1}\mid\omega_{i_1j_1}\right]^2\right\}I_{\tau,i_1j_1}\overline{X}_{i_1j_1}\overline{X}'_{i_1j_1}.\end{aligned}$$

Given Assumption 4.1.4, a conditional version of Lyapunov's Central Limit Theorem holds, and hence

$$\Upsilon\left(\Omega_n\right)^{-1/2}\left\{\frac{1}{\sqrt{n(n-1)}}\sum_{i_1=1}^n\sum_{j_1\neq i_1}\xi_{i_1j_1,\tau}\right\}\Rightarrow\mathcal{N}\left(0,I\right).$$

Now, it follows from using 4.1.4 that $\|\Upsilon\left(\Omega_n\right)-\Upsilon_n\|\xrightarrow{p}0$ as $n\to\infty$. It follows then that the limiting distribution is independent of the conditionally values, and therefore, the limiting distribution continues to hold unconditionally, with Υ_n replacing $\Upsilon\left(\Omega_n\right)$. That is,

$$\Upsilon_n^{-1/2}\left\{\frac{1}{\sqrt{n(n-1)}}\sum_{i_1=1}^n\sum_{j_1\neq i_1}\xi_{i_1j_1,\tau}\right\}\Rightarrow\mathcal{N}\left(0,I\right).$$

Part 4: Limiting distribution of $\widehat{\theta}_n$

Consider the matrix Σ_n, defined as $\Sigma_n=\Gamma_0^{-1}\times\Upsilon_n\times\Gamma_0^{-1}$. The limiting distribution of the $\widehat{\theta}_n$

follows from the definitions of $\widehat{\Psi}_n^{-1}$ and Σ_n, and from applying Slutsky's theorem. In other words,

$$
\begin{aligned}
& \sqrt{n(n-1)}\Sigma_n^{-1/2}\left(\widehat{\theta}_n - \theta_0\right) \\
= \quad & \sqrt{n(n-1)}\Sigma_n^{-1/2} \times \left\{\widehat{\Gamma}_n^{-1}\left[\frac{1}{m_n}\sum_{\sigma\in\mathcal{N}_{m_n}}\left\{\tilde{W}_\sigma \tilde{D}_{\sigma,\tau}^* - E\left[\tilde{W}_\sigma \tilde{D}_\sigma^* \mid \Omega_\sigma\right]\right\}\right]\right\} \\
= \quad & \Gamma_0^{1/2} \times \Upsilon_n^{-1/2} \times \Gamma_0^{-1/2} \times \left\{\frac{1}{\sqrt{n(n-1)}}\sum_{i_1=1}^{n}\sum_{j_1\neq i_1}\xi_{i_1 j_1,\tau}\right\} + o_p\left(1\right) \\
\Rightarrow \quad & \mathcal{N}\left(0, I\right).
\end{aligned}
$$

The proof is complete.

$\square$

B Technical Appendix

B.1 Equivalent representation for V statistics

The following lemma provides a U-statistic representation for a V-statistic when the kernel varies with n. Given n and for $m \leq n$, let $\sum_{(n,m)}$ denote the sum over the $\binom{n}{m}$ combinations of m distinct elements $(i_1, \cdots, i_m)$ from $(1, \cdots, n)$, and $\sum_{\Pi_{m!}}$ denotes the sum over the $m!$ permutations $(i_1, \cdots, i_m)$ of $(1, \cdots, m)$.

Let V_n be a V-statistic or order m, without common indices

$$V_n = \frac{1}{n^m} \sum_{i_1, \cdots, i_m = 1}^{n} \frac{1}{h^L} \gamma(X_{i_1}, \cdots, X_{i_m}) \mathbf{1} [i_1 \neq \cdots \neq i_m]$$

where $h \to 0$ as $n \to \infty$, and $\gamma : \mathbb{R}^L \mapsto \mathbb{R}$.

Let

$$\begin{aligned} U_n &= \binom{n}{m}^{-1} \sum_{(n,m)} \phi_h(X_{i_1}, \cdots, X_{i_m}) \\ \phi_h(X_1, \cdots, X_m) &= \frac{1}{m!} \sum_{\Pi_{m!}} \frac{1}{h^L} \gamma(X_{\pi_1}, \cdots, X_{\pi_m}) \end{aligned}$$

Lemma B.1. *Suppose that $E \parallel \gamma(X_{i_1}, \cdots, X_{i_m}) \parallel^2 < \infty$ for all $1 \leq i_1, \cdots, i_m \leq m$ and $m \leq n$, and $nh^2 \to \infty$. Then,*

$$V_n - U_n = o_p(1).$$

Proof. Let

$$\gamma_h(X_{i_1}, \cdots, X_{i_m}) = \frac{1}{h^L} \gamma(X_{i_1}, \cdots, X_{i_m}),$$

and notice that

$$\begin{aligned} n^m V_n &= \sum_{(n,m)} \sum_{\Pi_{m!}} \gamma_h(X_{\pi_1}, \cdots, X_{\pi_m}) \qquad (13) \\ &= [n(n-1) \cdots (n-m+1)] \binom{n}{m}^{-1} \sum_{(n,m)} \phi_h(X_{i_1}, \cdots, X_{i_m}) \\ &= [n(n-1) \cdots (n-m+1)] \, U_n, \end{aligned}$$

and hence, $(U_n - V_n) = O(n^{-1}) U_n$.

Consider now

$$E \left[(U_n - V_n)^2 \right] = O \left(\frac{1}{n^2} \right) E \left[U_n^2 \right],$$

and notice that a Cauchy-Schwarz inequality yields

$$
\begin{aligned}
E\left[U_n^2\right] &= \binom{n}{m}^{-2} E\left[\left(\sum_{(n,m)} \phi_h(X_{i_1}, \cdots, X_{i_m})\right)^2\right] \\
&\leq \binom{n}{m}^{-2} \binom{n}{m}^{2} E\left[\phi_h(X_{i_1}, \cdots, X_{i_m})^2\right]
\end{aligned}
$$

where

$$
\begin{aligned}
E\left[\phi_h(X_{i_1}, \cdots, X_{i_m})^2\right] &= \frac{1}{h^{2L}} O\left(E\left[\gamma(X_{i_1}, \cdots, X_{i_m})^2\right]\right) \\
&= O\left(\frac{1}{h^{2L}}\right)
\end{aligned}
$$

since $E \parallel \gamma(X_{i_1}, \cdots, X_{i_m}) \parallel^2 < \infty$ by assumption, and hence,

$$
E\left[(U_n - V_n)^2\right] \leq O\left(\frac{1}{(nh^L)^2}\right) = o(1)
$$

as $nh^L \to \infty$.

□

Notice that, unlike Lemma 5.7.3 in Serfling (2009, page 206) and Theorem 1 in Lee (2019, page 183), in equation 13 the average of terms with at least one common index is equal to zero due to the specification of the V-statistic without common indices.

B.2 Consistency for V-statistics

Lemma B.2. *Suppose that the Assumptions in Theorem 4.1 hold. Then*

$$
\begin{aligned}
\frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} \tilde{W}_\sigma \varphi_{l_1 l_2, \tau} \left\{ \frac{\hat{f}_{x, l_1 l_2}}{f_{vx, l_1 l_2}} \right\} - E\left[\tilde{W}_\sigma \varphi_{l_1 l_2, \tau} \frac{f_{x, l_1 l_2}}{f_{vx, l_1 l_2}}\right] &= o_p(1) \\
\frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} \tilde{W}_\sigma \varphi_{l_1 l_2, \tau} \left\{ \frac{f_{x, l_1 l_2}}{f_{vx, l_1 l_2}} \left(\frac{\hat{f}_{vx, l_1 l_2}}{f_{vx, l_1 l_2}}\right) \right\} - E\left[\tilde{W}_\sigma \varphi_{l_1 l_2, \tau} \frac{f_{x, l_1 l_2}}{f_{vx, l_1 l_2}}\right] &= o_p(1)
\end{aligned}
$$

with $(l_1, l_2) \in \{(i_1, j_1), (i_1, j_2), (i_2, j_1), (i_2, j_2)\}$ *for a given tetrad* $\sigma\{i_1, j_1; i_2, j_2\} \in \mathcal{N}_{m_n}$.

Proof. This proof focuses on the first result since the second one follows from similar arguments. Let

$$
\hat{V}_n = \frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} \tilde{W}_\sigma \frac{\varphi_{l_1 l_2, \tau}}{f_{vx, l_1 l_2}} \hat{f}_{x, l_1 l_2},
$$

and recall that the kernel estimator $\hat{f}_{x,l_1l_2}$ is defined as

$$\hat{f}_{x,l_1l_2} = \frac{1}{(n-2)(n-3)} \sum_{k_1 \neq i_1, j_1} \sum_{k_2 \neq i_1, j_1, k_1} \frac{1}{h^L} K_{x,h}\left(X_{k_1} - X_{l_1}, X_{k_2} - X_{l_2}\right).$$

Plugging in $\hat{f}_{x,l_1l_2}$ into $\hat{V}_n$ yields the following V-statistic of order six

$$6! \binom{n}{6}^{-1} \sum_{i_1 \neq i_2 \neq j_1 \neq j_2 \neq k_1 \neq k_2} \frac{1}{h^L} \tilde{W}_{i_1 i_2; j_1 j_2} \frac{\varphi_{l_1 l_2, \tau}}{f_{vx, l_1 l_2}} K_{x,h}\left(X_{k_1} - X_{l_1}, X_{k_2} - X_{l_2}\right).$$

Assumptions 4.1.1 and 4.1.5 imply that

$$E\left[|| \frac{1}{h^L} \tilde{W}_{i_1 i_2; j_1 j_2} \frac{\varphi_{l_1 l_2, \tau}}{f_{vx, l_1 l_2}} K_{x,h}\left(X_{k_1} - X_{l_1}, X_{k_2} - X_{l_2}\right) ||^2 \right] < \infty,$$

it follows then from Lemma B.1 that $\hat{V}_n$ is asymptotically equivalent to a six-order U-statistic as $nh^L \to \infty$. In particular, $\left(U_n - \hat{V}_n\right) = o_p(1)$ where

$$\begin{aligned} U_n &= \binom{n}{6}^{-1} \sum_{i_1 < \cdots < i_6} \phi_{\bar{\sigma}\{i_1, \cdots, i_6\}, \tau} \\ \phi_{\bar{\sigma}\{i_1, \cdots, i_6\}, \tau}, &= (6!)^{-1} \sum_{\pi \in \Pi_{6!}} \frac{1}{h^L} \tilde{W}_{\pi_1 \pi_2; \pi_3 \pi_4} \frac{\varphi_{\pi_{l_1} \pi_{l_2}, \tau}}{f_{vx, \pi_{l_1} \pi_{l_2}}} K_{x,h}\left(X_{\pi_5} - X_{\pi_{l_1}}, X_{\pi_6} - X_{\pi_{l_2}}\right) \end{aligned}$$

where $\sum_{i_1 < \cdots < i_6}$ denotes sum over the $\binom{n}{6}$ combinations of 6 distinct elements $(i_1, \cdots, i_6)$ from $(1, \cdots, n)$, and $\bar{\sigma}\{i_1, \cdots, i_6\}$ is used to denote the 6-tuple $\{i_1, \cdots, i_6\}$.

U_n is a sixth order U-statistic where the kernel $\phi_{\bar{\sigma}, \tau}$ varies with n as in Powell, Stock, and Stoker (1989). Using Lemma A.3 in Ahn and Powell (1993), it is sufficient to show $E\left[|| \phi_{\bar{\sigma}, \tau} ||^2\right] = o(n)$ to conclude that $U_n - E\left[\frac{1}{h^L} \tilde{W}_{i_1 i_2; j_1 j_2} \frac{\varphi_{l_1 l_2, \tau}}{f_{vx, l_1 l_2}} K_{x,h}\left(X_{k_1} - X_{l_1}, X_{k_2} - X_{l_2}\right)\right] = o_p(1)$.

A Cauchy-Schwarz inequality can be used to show that the expectation

$$E\left[|| \frac{1}{6!} \sum_{\pi \in \Pi_{6!}} \frac{1}{h^L} \tilde{W}_{\pi_1 \pi_2; \pi_3 \pi_4} \frac{\varphi_{\pi_{l_1} \pi_{l_2}, \tau}}{f_{vx, \pi_{l_1} \pi_{l_2}}} K_{x,h}\left(X_{\pi_5} - X_{\pi_{l_1}}, X_{\pi_6} - X_{\pi_{l_2}}\right) ||^2 \right]$$

is bounded above by

$$\frac{1}{6! h^{2L}} \sum_{\pi \in \Pi_{6!}} E\left[\left(\frac{\varphi_{\pi_{l_1} \pi_{l_2}, \tau}}{f_{vx, \pi_{l_1} \pi_{l_2}}}\right)^2 K_{x,h}\left(X_{\pi_5} - X_{\pi_{l_1}}, X_{\pi_6} - X_{\pi_{l_2}}\right)^2 \tilde{W}_{\pi_1 \pi_2; \pi_3 \pi_4} \tilde{W}'_{\pi_1 \pi_2; \pi_3 \pi_4} \right].$$

Let $X_{\bar{\sigma}\{\pi_1,\cdots,\pi_6\}} = \{X_{\pi_1},\cdots,X_{\pi_6}\}$, and observe that

$$
\begin{aligned}
& E\left[\frac{1}{h^{2L}}\left(\frac{\varphi_{\pi_{l_1}\pi_{l_2},\tau}}{f_{vx,\pi_{l_1}\pi_{l_2}}}\right)^2 K_{x,h}\left(X_{\pi_5}-X_{\pi_{l_1}},X_{\pi_6}-X_{\pi_{l_2}}\right)^2 \tilde{W}_{\pi_1\pi_2;\pi_3\pi_4}\tilde{W}'_{\pi_1\pi_2;\pi_3\pi_4}\right] \\
=& E\left[E\left[\left(\frac{\varphi_{\pi_{l_1}\pi_{l_2},\tau}}{f_{vx,\pi_{l_1}\pi_{l_2}}}\right)^2 \mid X_{\bar{\sigma}\{\pi_1,\cdots,\pi_6\}}\right]\frac{1}{h^{2L}}K_{x,h}\left(X_{\pi_5}-X_{\pi_{l_1}},X_{\pi_6}-X_{\pi_{l_2}}\right)^2 \tilde{W}_{\pi_1\pi_2;\pi_3\pi_4}\tilde{W}'_{\pi_1\pi_2;\pi_3\pi_4}\right] \\
\leq& \frac{1}{h^{2L}}\sup_{i_1i_2;i_3i_4}\left(\tilde{W}_{i_1i_2;i_3i_4}\right)\sup_{i_1i_2;i_3i_4}\left(\tilde{W}_{i_1i_2;i_3i_4}\right)' \\
& \times E\left[E\left[\left(\frac{\varphi_{\pi_{l_1}\pi_{l_2},\tau}}{f_{vx,\pi_{l_1}\pi_{l_2}}}\right)^2 \mid X_{\pi_{l_1}},X_{\pi_{l_2}}\right]K_{x,h}\left(X_{\pi_5}-X_{\pi_{l_1}},X_{\pi_6}-X_{\pi_{l_2}}\right)^2\right] \\
\leq& \frac{1}{h^{2L}}\sup_{i_1i_2;i_3i_4}\left(\tilde{W}_{i_1i_2;i_3i_4}\right)\sup_{i_1i_2;i_3i_4}\left(\tilde{W}_{i_1i_2;i_3i_4}\right)'\sup_{(x,x')\in\mathbb{S}_x,\tau\geq 0}\left(E\left[\left(\frac{\varphi_{i_1i_2,\tau}}{f_{vx,l_1l_2}}\right)^2 \mid X_{\pi_{l_1}},X_{\pi_{l_2}}\right]\right) \\
& \times E\left[K_{x,h}\left(X_{\pi_5}-X_{\pi_{l_1}},X_{\pi_6}-X_{\pi_{l_2}}\right)^2\right] \\
=& O\left(\frac{1}{h^L}\right)\times\int K_x\left[\nu_1,\nu_2\right]^2 f(X_{l_1},X_{l_2})f(X_{l_1}+\nu_1h,X_{l_2}+\nu_2h)dX_{l_1}dX_{l_2}d\nu_1d\nu_2 \\
=& h^{-L}O(1) = O(n(nh^L)^{-1}) = o(n),
\end{aligned}
$$

where the first inequality follows from Assumptions 3.1.1 and 4.1.1. The second inequality follows from Assumption 4.1.4. The second to last equality follows from Assumption 4.1.1, and the change of variables $X_{i_5} = X_{l_1}+\nu_1h$ and $X_{i_6} = X_{l_2}+\nu_2h$ with Jacobian h^L. The last equality follows from Assumption 4.1.5.

Consequently, $E\left[\|\ \phi_{\bar{\sigma},\tau}\ \|^2\right] = o(n)$ if $nh^L\to\infty$. Thus, Lemma A.3 in Ahn and Powell (1993) implies that

$$
U_n - E\left[\frac{1}{h^L}\tilde{W}_{\pi_1\pi_2;\pi_3\pi_4}\left\{\frac{\varphi_{\pi_{l_1}\pi_{l_2},\tau}}{f_{vx,\pi_{l_1}\pi_{l_2}}}\right\}K_{x,h}\left(X_{\pi_5}-X_{\pi_{l_1}},X_{\pi_6}-X_{\pi_{l_2}}\right)\right] \quad = \quad o_p(1)
$$

as $n\to\infty$.

Notice that

$$
\begin{aligned}
& E\left[\frac{1}{h^L}\tilde{W}_{i_1i_2;i_3i_4}\left\{\frac{\varphi_{l_1l_2,\tau}}{f_{vx,l_1l_2}}\right\}K_{x,h}\left(X_{i_5}-X_{l_1},X_{i_6}-X_{l_2}\right)\right] \\
=& \frac{1}{h^L}E\left[E\left[\tilde{W}_{i_1i_2;i_3i_4}\left\{\frac{\varphi_{l_1l_2,\tau}}{f_{vx,l_1l_2}}\right\}\mid X_{\bar{\sigma}\{\pi_1,\cdots,\pi_6\}}\right]K_{x,h}\left[X_{i_5}-X_{l_1},X_{i_6}-X_{l_2}\right]\right] \\
=& \frac{1}{h^L}E\left[E\left[\tilde{W}_{i_1i_2;i_3i_4}\left\{\frac{\varphi_{l_1l_2,\tau}}{f_{vx,l_1l_2}}\right\}\mid X_{i_1},X_{i_2},X_{i_3},X_{i_4}\right]K_{x,h}\left[X_{i_5}-X_{l_1},X_{i_6}-X_{l_2}\right]\right] \\
=& \int E\left[\tilde{W}_{i_1i_2;i_3i_4}\left\{\frac{\varphi_{l_1l_2,\tau}}{f_{vx,l_1l_2}}\right\}\mid X_{i_1},X_{i_2},X_{i_3},X_{i_4}\right] \\
& \qquad \times K_{x,h}\left[X_{i_5}-X_{l_1},X_{i_6}-X_{l_2}\right]f(X_{\bar{\sigma}\{i_1,\cdots,i_6\}})dX_{\bar{\sigma}\{i_1,\cdots,i_6\}}
\end{aligned}
$$

where the second equality follows from Assumption 3.1.1 Next, consider the change of variables $X_{i_5} = X_{l_1} + h\nu_1$ and $X_{i_6} = X_{l_2} + h\nu_2$ with Jacobian h^L. It follows then

$$\int E\left[\tilde{W}_{i_1 i_2; i_3 i_4}\left\{\frac{\varphi_{l_1 l_2,\tau}}{f_{vx,l_1 l_2}}\right\} \mid X_{i_1}, X_{i_2}, X_{i_3}, X_{i_4}\right] \\ \times K\left[\nu_1, \nu_2\right] f(X_{i_1}, \cdots, X_{i_4})\left\{f(X_{l_1} + h\nu_1, X_{l_2} + h\nu_2)\right\} dX_{i_1}\cdots, dX_{i_4} d\nu_1 d\nu_2.$$

Assumption 4.1.1 guarantees that $f_x(\cdot,\cdot)$ is $\overline{M}$-times differentiable with respect to all of its arguments, and Assumption 4.1.5 ensures that $K_x(\cdot,\cdot)$ is a bias-reducing kernel of order $2\overline{M}$. It follows from an $\overline{M}$-order Taylor expansion $f(X_{l_1} + h\nu_1, X_{l_2} + h\nu_2)$ around $f(X_{i_1}, X_{i_3})$, and the properties of the kernel that

$$\begin{aligned} & \int E\left[\tilde{W}_{i_1 i_2; i_3 i_4}\left\{\frac{\varphi_{l_1 l_2,\tau}}{f_{vx,l_1 l_2}}\right\} f_{x,l_1 l_2} \mid X_{i_1}, X_{i_2}, X_{i_3}, X_{i_4}\right] f(X_{i_1}, \cdots, X_{i_4}) dX_{i_1}\cdots, X_{i_4} + h^{\overline{M}} O(1) \\ = \;& E\left[\tilde{W}_{i_1 i_2; i_3 i_4}\varphi_{l_1 l_2,\tau}\left\{\frac{f_{x,l_1 l_2}}{f_{vx,l_1 l_2}}\right\}\right] + h^{\overline{M}} O(1) \\ = \;& E\left[\tilde{W}_{i_1 i_2; i_3 i_4} D^*_{l_1 l_2,\tau}\right] + o(1). \end{aligned}$$

The proof is complete. □

B.3 Lemmas for Asymptotic Normality Theorem

Notation

The following notation will prove to be useful to show Lemmas B.3-B.6. For any finite n, let $\Omega_n = \{X_n, A_n, v_n\}$. Given a fix tetrad $\sigma\{i_1, i_2; j_1, j_2\} \in \mathcal{N}_{m_n}$, let

$$X_\sigma = \{X_{i_1}, X_{i_2}, X_{j_1}, X_{j_2}\}, \quad A_\sigma = \{A_{i_1}, A_{i_2}, A_{j_1}, A_{j_2}\}, \quad v_\sigma = \{v_{i_1}, v_{i_2}, v_{j_1}, v_{j_2}\}, \quad \Omega_\sigma = \{X_\sigma, A_\sigma, v_\sigma\},$$

and for any dyad $(l_1, l_2) \in \{(i_1, j_1), (i_1, j_2), (i_2, j_1), (i_2, j_2)\}$, define

$$\begin{aligned} \omega_{l_1 l_2} &= \{X_{l_1}, X_{l_2}, A_{l_1}, A_{l_2}, v_{l_1 l_2}\} \\ T^\dagger_{l_1 l_2} &= T_{l_1 l_2} - E\left[\tilde{W}_\sigma \tilde{D}^*_{\sigma,\tau} \mid \Omega_\sigma\right] \end{aligned}$$

for any random variable $T_{l_1 l_2}$.

Lemma B.3. *Suppose that the Assumptions in Theorem 4.2 hold, and consider*

$$\widehat{\eta}_{[l_1 l_2],\tau} = \frac{1}{m_n}\sum_{\sigma\in\mathcal{N}_{m_n}} \tilde{W}_\sigma \varphi_{l_1 l_2,\tau}\left(\frac{\widehat{f}_{x,l_1 l_2}}{\widehat{f}_{vx,\sigma l_1 l_2}}\right).$$

with $(l_1, l_2) \in \{(i_1, j_1), (i_1, j_2), (i_2, j_1), (i_2, j_2)\}$. *It follows that* $\widehat{\eta}_{[l_1 l_2],\tau}$ *can be written as*

$$\widehat{\eta}_{[l_1 l_2],\tau} = \frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} \tilde{W}_\sigma \varphi_{l_1 l_2,\tau} \left\{ \frac{f_{x,l_1 l_2}}{f_{vx,l_1 l_2}} + \frac{\widehat{f}_{x,l_1 l_2} - f_{x,l_1 l_2}}{f_{vx,l_1 l_2}} - \frac{f_{x,l_1 l_2}}{f_{vx,l_1 l_2}} \times \frac{\widehat{f}_{vx,l_1 l_2} - f_{vx,l_1 l_2}}{f_{vx,l_1 l_2}} \right\} + o_p(1).$$

Proof. Given $h \to 0$,and $n^{1-\delta} h^{L+1} \to \infty$ for any $\delta > 0$, it follows from a variance calculation argument that

$$\sup_{(v,x,x') \in \overline{\Omega}_{v,x}} \mid \hat{f}_{vx}(v,x,x') - f_{vx}(v,x,x') \mid = o_p\left(1\right)$$
$$\sup_{(x,x') \in \overline{\Omega}_{x}} \mid \hat{f}_{x}(x,x') - f_{x}(x,x') \mid = o_p\left(1\right),$$

for any $\delta > 0$. See, e.g., Silverman (1978), Collomb and Härdle (1986),Aradillas-Lopez (2010), and for applications to network models Leung (2015b) and Graham et al. (2019).

Consider a second order Taylor expansion of $\widehat{f}_{x,l_1 l_2}/\widehat{f}_{vx,l_1 l_2}$ around $f_{x,l_1 l_2}/f_{vx,l_1 l_2}$. The quadratic terms in the expansion involve second order derivatives of $f_{x,l_1 l_2}/f_{vx,l_1 l_2}$ evaluated at $\tilde{f}_{x,l_1 l_2}$ and $\tilde{f}_{vx,l_1 l_2}$, where $\tilde{f}_{x,l_1 l_2}$ lies in between $\widehat{f}_{x,l_1 l_2}$ and $f_{x,l_1 l_2}$, and similarly $\tilde{f}_{vx,l_1 l_2}$ lies in between $\widehat{f}_{vx,l_1 l_2}$ and $f_{vx,l_1 l_2}$. From substituting a second order Taylor expansion of $\widehat{f}_{x,l_1 l_2}/\widehat{f}_{vx,l_1 l_2}$ around $f_{x,l_1 l_2}/f_{vx,l_1 l_2}$ into $\widehat{\eta}_{[l_1 l_2],\tau}$, I obtain

$$\widehat{\eta}_{[l_1 l_2],\tau} = \frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} \tilde{W}_\sigma \varphi_{l_1 l_2,\tau} \left\{ \frac{f_{x,l_1 l_2}}{f_{vx,l_1 l_2}} + \frac{\widehat{f}_{x,l_1 l_2} - f_{x,l_1 l_2}}{f_{vx,l_1 l_2}} - \frac{f_{x,l_1 l_2}}{f_{vx,l_1 l_2}} \times \frac{\widehat{f}_{vx,l_1 l_2} - f_{vx,l_1 l_2}}{f_{vx,l_1 l_2}} \right\} + R_n,$$

where R_n denotes the reminder term. The result follows from showing that $R_n = o_p(1)$.

The first component of R_n is

$$\begin{aligned}
& \frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} \tilde{W}_\sigma \varphi_{l_1 l_2,\tau} \left\{ \tilde{f}_{x,l_1 l_2} \frac{\left(\widehat{f}_{vx,l_1 l_2} - f_{vx,l_1 l_2}\right)^2}{\tilde{f}^3_{vx,l_1 l_2}} \right\} \\
\leq & \left[\sup_{(x,x') \in \overline{\Omega}_x} |f_x| \right] \left[\sup_{(v,x,x') \in \overline{\Omega}_{vx}} |f_{vx}^{-3}| \right] \left[\sup_{(v,x,x') \in \overline{\Omega}_{vx}} |\widehat{f}_{vx} - f_{vx}| \right]^2 \left(\frac{1}{m_n} \sum_{\sigma \in \mathcal{N}_{m_n}} || \tilde{W}_\sigma \varphi_{l_1 l_2,\tau} || \right) \\
= & O_p(1) \left[\sup_{(v,x,x')} |\widehat{f}_{vx} - f_{vx}| \right]^2 \\
= & o_p(1).
\end{aligned}$$

The first inequality follows from Assumption 4.1.1. The equality follows from the fact that the V-statistic inside the parenthesis converges to its expectation given that Assumptions 3.1.1 and 4.1.1. The result follows from the uniform convergence of the kernel estimator.

The remaining component of R_n is

$$\begin{aligned}
& \frac{1}{m_n}\sum_{\sigma\in\mathcal{N}_{m_n}}\tilde{W}_\sigma\varphi_{l_1l_2,\tau}\left\{\frac{(\widehat{f}_{vx,l_1l_2}-f_{vx,l_1l_2})(\widehat{f}_{x,l_1l_2}-f_{x,l_1l_2})}{f^2_{vx,l_1l_2}}\right\} \\
\leq & \left[\sup_{(v,x,x)\in\overline{\Omega}_{vx}}\mid f_{vx}^{-2}\mid\right]\left[\sup_{(v,x,x)\in\overline{\Omega}_{vx}}\mid \widehat{f}_{vx}-f_{vx}\mid\right]\left[\sup_{(x,x)\in\overline{\Omega}_{x}}\mid \widehat{f}_{x}-f_{x}\mid\right] \\
& \times\left(\frac{1}{m_n}\sum_{\sigma\in\mathcal{N}_{m_n}}\|\tilde{W}_\sigma\varphi_{l_1l_2,\tau}\|\right) \\
= & O_p(1)\left[\sup_{(v,x,x)\in\overline{\Omega}_{vx}}\mid \widehat{f}_{vx}-f_{vx}\mid\right]\left[\sup_{(x,x)\in\overline{\Omega}_{vx}}\mid \widehat{f}_{x}-f_{x}\mid\right]. \\
= & o_p(1).
\end{aligned}$$

The result follows from the uniform convergence of the kernel estimators. This completes the proof.

□

Lemma B.4. *Under the same Assumptions of Theorem 4.2, it follows that the Hájek projection of*

$$\begin{aligned}
S^\dagger_{1,n\tau} &= S_{1,n\tau}-E\left[\tilde{W}_\sigma\tilde{D}^*_{\sigma,\tau}\mid\Omega_n\right] \\
&= \frac{1}{m_n}\sum_{\sigma\in\mathcal{N}_{m_n}}\left\{\tilde{W}_\sigma\tilde{D}^*_{\sigma,\tau}-E\left[\tilde{W}_\sigma\tilde{D}^*_{\sigma,\tau}\mid\Omega_\sigma\right]\right\}
\end{aligned}$$

into an arbitrary function $\zeta_{i_1j_1}=(X_{i_1},X_{j_1},A_{i_1},A_{j_1},v_{i_1j_1},U_{i_1j_1})$ *is given by*

$$V^*_{1,n\tau} = \frac{1}{n(n-1)}\sum_{i_1=1}^n\sum_{j_1\neq i_1}\xi_{i_1j_1,\tau}$$

and

$$n(n-1)\Upsilon_n^{-1/2}E\left[\left(S^\dagger_{1,n\tau}-V^*_{1,n\tau}\right)^2\right]\Upsilon_n^{-1/2}=o(1),$$

where $\Upsilon_n=n(n-1)Var(V^*_{1,n\tau})$ *and* $Var(V^*_{1,n\tau})=O_p(p_n^2\tau^2)$.

Proof. **Step 1. Hájek Projection**

Consider the tetrad $\sigma\{i_1,i_2;j_1,j_2\}$, let

$$\begin{aligned}
s\left(\sigma\{i_1,i_2;j_1,j_2\}\right) &= \tilde{W}_\sigma\tilde{D}^*_{\sigma,\tau}-E\left[\tilde{W}_\sigma\tilde{D}^*_{\sigma,\tau}\mid\Omega_\sigma\right] \\
&= \tilde{W}_\sigma\left\{\tilde{D}^*_{\sigma,\tau}-E\left[\tilde{D}^*_{\sigma,\tau}\mid\Omega_\sigma\right]\right\},
\end{aligned}$$

and notice that

$$\begin{aligned} E\left[s\left(\sigma\{i_1,i_2;j_1,j_2\}\right) \mid \zeta_{i_1j_1}\right] &= E\left[\tilde{W}_\sigma\left\{\tilde{D}^*_{\sigma,\tau} - E\left[\tilde{D}^*_{\sigma,\tau} \mid \Omega_\sigma\right]\right\} \mid \zeta_{i_1j_1}\right] \\ &= \left\{D^*_{i_1j_1,\tau} - E\left[D^*_{i_1j_1,\tau} \mid \omega_{i_1j_1}\right]\right\} E\left[\tilde{W}_\sigma \mid X_{i_1j_1}\right]. \end{aligned}$$

where the second equality follows from the Law of Iterated Expectations, and Assumptions 3.1.1 and 3.1.2. To be precise, observe that for $\{l_1,l_2\} \neq \{i_1,j_1\}$ with $(l_1,l_2) \in \{(i_1,j_2),(i_2,j_1),(i_2,j_2)\}$,

$$\begin{aligned} & E\left[\tilde{W}_\sigma\left\{\tilde{D}^*_{l_1l_2,\tau} - E\left[\tilde{D}^*_{l_1l_2,\tau} \mid \Omega_\sigma\right]\right\} \mid \zeta_{i_1j_1}\right] \\ = \; & E\left[\tilde{W}_\sigma\left\{E\left[\tilde{D}^*_{l_1l_2,\tau} \mid \omega_{l_1l_2}\right] - E\left[\tilde{D}^*_{l_1l_2,\tau} \mid \omega_{l_1l_2}\right]\right\} \mid \zeta_{i_1j_1}\right] \\ = \; & 0. \end{aligned}$$

It follows then that the Hájek projection is given by

$$V^*_{1,n\tau} = \frac{1}{n(n-1)} \sum_{i_1=1}^{n} \sum_{j_1 \neq i_1} \xi_{i_1j_1,\tau},$$

with

$$\begin{aligned} \xi_{i_1j_1,\tau} &= \left\{D^*_{i_1j_1} - E\left[D^*_{i_1j_1} \mid \omega_{i_1j_1}\right]\right\} I_{\tau,i_1j_1} \overline{\chi}_{i_1j_1} \\ \overline{\chi}_{i_1j_1} &= \left\{\frac{1}{(n-2)(n-3)} \sum_{i_2 \neq i_1,j_1} \sum_{j_2 \neq i_1,j_1,i_2} E\left[\tilde{W}_{\sigma\{i_1,i_2;j_1,j_2\}} \mid X_{i_1}, X_{j_1}\right]\right\}. \end{aligned}$$

Notice that $E\left[V^*_{1,n\tau}\right] = E\left[\xi_{i_1j_1,\tau}\right] = 0.$

Step 2. Variance of Hájek Projection

For two different dyads $\{i_1,j_1\} \neq \{i'_1,j'_1\}$ with zero common indices, Assumption 3.1.1 implies that

$$E\left[\xi_{i_1j_1,\tau}\xi_{i'_1j'_1,\tau}\right] = E\left[\xi_{i_1j_1,\tau}\right] E\left[\xi_{i'_1j'_1,\tau}\right] = 0.$$

Observe that for two dyads $\{i_1,j_1\} \neq \{i_1,j'_1\}$ with one common index, the conditionally independent formation of links implies by Assumption 3.1.2 yields

$$E\left[\xi_{i_1j_1,\tau}\xi_{i'_1j'_1,\tau}\right] = E\left[E\left[\xi_{i_1j_1,\tau} \mid \Omega_n\right] E\left[\xi_{i'_1j'_1,\tau} \mid \Omega_n\right]\right] = 0.$$

Therefore, the variance of $V_{1,n\tau}^*$ is given by

$$\begin{aligned}
Var\left(V_{1,n\tau}^*\right) &= \left\{\frac{1}{n(n-1)}\right\}^2 \left\{\sum_{i_1=1}^{n}\sum_{j_1\neq i_1} E\left[\xi_{i_1 j_1,\tau}\xi'_{i'_1 j'_1,\tau}\right]\right\} \\
&= \left\{\frac{1}{n(n-1)}\right\}^2 \left\{\sum_{i_1=1}^{n}\sum_{j_1\neq i_1} \Lambda^*_{i_1,j_1}\right\}
\end{aligned}$$

where

$$\Lambda^*_{i_1,j_1} = E\left[\left\{E\left[D^*_{i_1 j_1}D^*_{i_1 j_1} \mid \omega_{i_1 j_1}\right] - E\left[D^*_{i_1 j_1} \mid \omega_{i_1 j_1}\right]^2\right\} I^2_{\tau,i_1 j_1}\overline{\chi}_{i_1 j_1}\overline{\chi}'_{i_1 j_1}\right].$$

Define

$$\Upsilon_{n,\tau} = n(n-1)Var\left(V_{1,n\tau}^*\right) = \frac{1}{n(n-1)}\left\{\sum_{i_1=1}^{n}\sum_{j_1\neq i_1} \Lambda^*_{i_1,j_1}\right\}.$$

Step 3. Variance of $S_{1,n\tau}^{\dagger}$

Given two different tetrads $\sigma\{i_1,i_2;j_1,j_2\}$ and $\sigma'\{i'_1,i'_2;j'_1,j'_2\}$, let

$$\Delta_{c,n} = Cov\left(s\left(\sigma\{i_1,i_2;j_1,j_2\}\right), s\left(\sigma'\{i'_1,i'_2;j'_1,j'_2\}\right)\right)$$

denote the covariance between $s(\sigma)$ and $s(\sigma')$ when $\sigma\{i_1,i_2;j_1,j_2\}$ and $\sigma'\{i'_1,i'_2;j'_1,j'_2\}$ have $c = 0,1,2,3,4$ indices in common.

It follows from conditionally independence formation of links, implied by Assumption 3.1.2, and the conditional mean zero, $E\left[s\left(\sigma\{i_1,i_2;j_1,j_2\}\right) \mid \Omega_\sigma\right] = 0$, that $\Delta_{0,n} = \Delta_{1,n} = 0$.

Consider

$$\begin{aligned}
\Delta_{2,n} &= E\left[s(\sigma\{i_1,i_2;j_1,j_2\})s(\sigma'\{i_1,i'_2;j_1,j'_2\})'\right] \\
&= E\left[\left\{\tilde{D}^*_{\sigma,\tau} - E\left[\tilde{D}^*_{\sigma,\tau} \mid \Omega_\sigma\right]\right\}\left\{\tilde{D}^*_{\sigma',\tau} - E\left[\tilde{D}^*_{\sigma',\tau} \mid \Omega_{\sigma'}\right]\right\}\tilde{W}_\sigma\tilde{W}_{\sigma'}\right] \\
&= E\left[\left\{E\left[\tilde{D}^*_{i_1 j_1,\tau}\tilde{D}^*_{i_1 j_1,\tau} \mid \omega_{i_1 j_1}\right] - E\left[\tilde{D}^*_{i_1 j_1,\tau} \mid \omega_{i_1 j_1}\right]^2\right\} I^2_{\tau,i_1 j_1}\tilde{W}_\sigma\tilde{W}_{\sigma'}\right].
\end{aligned}$$

It follows from the results above that $Var\left(S^{\dagger}_{1,nt}\right)$ can be expanded as

$$
\begin{aligned}
& Var\left(S^{\dagger}_{1,nt}\right) \\
= & \left(\frac{1}{m_n}\right)^2 \sum_{\sigma\in\mathcal{N}_{m_n}} \sum_{\sigma'\in\mathcal{N}_{m_n}} \left\{E\left[s(\sigma\{i_1,i_2;j_1,j_2\})s(\sigma'\{i_1,i_2';j_1',j_2'\})'\right]\right\} \\
= & \left(\frac{1}{m_n}\right)^2 \sum_{i_1=1}^{n} \sum_{j_1\neq i_1} \left\{\sum_{k_1\neq i_1,j_1} \sum_{k_2\neq i_1,j_1,k_1} \sum_{l_1\neq i_1,j_1} \sum_{l_2\neq i_1,j_1,l_1} \Delta_{2,n}\right\} + O\left(\frac{\Delta_{3,n}}{n^3}\right) + O\left(\frac{\Delta_{4,n}}{n^4}\right).
\end{aligned}
$$

Notice that the term inside the brackets scaled by $[(n-2)(n-3)]^{-2}$ is equivalent to $\Lambda^*_{i_1j_1}$, in particular,

$$
\begin{aligned}
\Lambda^*_{i_1j_1} &= \left\{\frac{1}{(n-2)(n-3)}\right\}^2 \sum_{k_1\neq i_1,j_1} \sum_{k_2\neq i_1,j_1,k_1} \sum_{l_1\neq i_1,j_1'} \sum_{l_2\neq i_1,j_1',l_1} \Delta_{2,n} \\
&= E\left[\left\{E\left[\tilde{D}^*_{i_1j_1,\tau}\tilde{D}^*_{i_1j_1,\tau} \mid \omega_{i_1j_1}\right] - E\left[\tilde{D}^*_{i_1j_1,\tau} \mid \omega_{i_1j_1}\right]^2\right\} I^2_{\tau,i_1j_1}\overline{\chi}_{i_1j_1}\overline{\chi}'_{i_1j_1}\right],
\end{aligned}
$$

which follows from the definition of $\overline{\chi}_{i_1j_1}$.

Hence,

$$
Var\left(S^{\dagger}_{1,nt}\right) = \left(\frac{1}{n(n-1)}\right)^2 \left\{\sum_{i_1=1}^{n} \sum_{j_1\neq i_1} \Lambda^*_{i_1,j_1}\right\} + o(1),
$$

and $Var\left(V^*_{1,n\tau}\right) - Var\left(S^{\dagger}_{1,n\tau}\right) = o(1)$.

Step 4. Asymptotic Equivalence

To show that

$$
n(n-1)\Upsilon^{-1/2}_{n,\tau} E\left[\left(S^{\dagger}_{1,n\tau} - V^*_{1,n\tau}\right)\left(S^{\dagger}_{1,n\tau} - V^*_{1,n\tau}\right)'\right]\Upsilon^{-1/2}_{n,\tau} = o(1)
$$

It is sufficient to prove that $Var\left(V^*_{1,n\tau}\right)^{-1/2} Cov\left[V^*_{1,n\tau}, S_{1,n\tau}\right] Var\left(V^*_{1,n\tau}\right)^{-1/2} = I$, which in turn, follows from noticing that

$$
\begin{aligned}
Cov\left[V^*_{1,n\tau}, S^{\dagger}_{1,n\tau}\right] &= E\left[V^*_{1,n\tau}, S^{\dagger}_{1,n\tau}\right] \\
&= E\left[V^*_{1,n\tau}\left(S^{\dagger}_{1,n\tau} - V^*_{1,n\tau}\right)'\right] + E\left[V^*_{1,n\tau}\left(V^*_{1,n\tau}\right)'\right] \\
&= Var(V^*_{1,n\tau}),
\end{aligned}
$$

since by construction of the orthogonal projection

$$
E\left[V^*_{1,n\tau}\left(S_{1,n\tau} - V^*_{1,n\tau}\right)'\right] = 0.
$$

The proof is complete. $\square$

Lemma B.5. *Under the same Assumptions of Theorem 4.2, it follows that the Hájek projection of*

$$
\begin{aligned}
S^{\dagger}_{2,n\tau} &= S_{2,n\tau} - E\left[\tilde{W}_\sigma \tilde{D}^*_{\sigma,\tau} \mid \Omega_\sigma\right] \\
S_{2,n\tau} &= \frac{1}{m_n}\sum_{\sigma\in\mathcal{N}_{m_n}} \tilde{W}_\sigma \left\{\left(\frac{\varphi_{i_1j_1,\tau}\widehat{f}_{x,i_1j_1}}{f_{vx,i_1j_1}} - \frac{\varphi_{i_1j_2,\tau}\widehat{f}_{x,i_1j_2}}{f_{vx,i_1j_2}}\right) - \left(\frac{\varphi_{i_2j_1,\tau}\widehat{f}_{x,i_2j_1}}{f_{vx,i_2j_1}} - \frac{\varphi_{i_2j_2,\tau}\widehat{f}_{x,i_2j_2}}{f_{vx,i_2j_2}}\right)\right\}
\end{aligned}
$$

into an arbitrary function $\zeta_{i_1j_1} = (X_{i_1}, X_{j_1}, A_{i_1}, A_{j_1}, v_{i_1j_1}, U_{i_1j_1})$ is given by

$$
V^*_{2,n\tau} = \frac{1}{n(n-1)}\sum_{i_1=1}^{n}\sum_{j_1\neq i_1}\bar{\xi}_{i_1j_1,\tau}
$$

and

$$
n\Upsilon_n^{-1/2} E\left[\left(S_{2,n\tau} - V^*_{2,n\tau}\right)^2\right]\Upsilon_n^{-1/2} = o(1),
$$

where $\Upsilon_n = nVar(V^*_{2,n\tau})$.

Proof. Similarly to the definition for tetrads, I introduce the function $\overline{\sigma} = \overline{\sigma}\{i_1, i_2; j_1, j_2; k_1, k_2\}$ that maps each unique 6-tuple $\{i_1, i_2; j_1, j_2; k_1, k_2\}$ into an index set $N_{\overline{m}_n} = \{1, \cdots, \overline{m}_n\}$ where $\overline{m}_n$ denotes the total number of those 6-tuples. Hence, each distinct 6-tuple $\{i_1, i_2; j_1, j_2; k_1, k_2\}$ corresponds to a unique $\overline{\sigma} = \overline{\sigma}\{i_1, i_2; j_1, j_2; k_1, k_2\} \in N_{\overline{m}_n}$.

Consider a fixed 6-tuple $\{i_1, i_2; j_1, j_2; k_1, k_2\}$, and define

$$
\begin{aligned}
s_{i_1,j_1}(\overline{\sigma}) &= \tilde{W}_{i_1i_2,j_1j_2}\left\{\frac{1}{h^L}\frac{\varphi_{i_1j_1,\tau}}{f_{vx,i_1j_1}}K_{x,h}\left(X_{k_1}-X_{i_1}, X_{k_2}-X_{j_1}\right) - E\left[D^*_{i_1j_1,\tau} \mid \Omega_{i_1i_2,j_1j_2}\right]\right\} \\
s_{i_1,j_2}(\overline{\sigma}) &= \tilde{W}_{i_1i_2,j_1j_2}\left\{\frac{1}{h^L}\frac{\varphi_{i_1j_2,\tau}}{f_{vx,i_1j_2}}K_{x,h}\left(X_{k_1}-X_{i_1}, X_{k_2}-X_{j_2}\right) - E\left[D^*_{i_1j_2,\tau} \mid \Omega_{i_1i_2,j_1j_2}\right]\right\} \\
s_{i_2,j_1}(\overline{\sigma}) &= \tilde{W}_{i_1i_2,j_1j_2}\left\{\frac{1}{h^L}\frac{\varphi_{i_2j_1,\tau}}{f_{vx,i_2j_1}}K_{x,h}\left(X_{k_1}-X_{i_2}, X_{k_2}-X_{j_1}\right) - E\left[D^*_{i_2j_1,\tau} \mid \Omega_{i_1i_2,j_1j_2}\right]\right\} \\
s_{i_2,j_2}\overline{\sigma}) &= \tilde{W}_{i_1i_2,j_1j_2}\left\{\frac{1}{h^L}\frac{\varphi_{i_2j_2,\tau}}{f_{vx,i_2j_2}}K_{x,h}\left(X_{k_1}-X_{i_2}, X_{k_2}-X_{j_2}\right) - E\left[D^*_{i_2j_2,\tau} \mid \Omega_{i_1i_2,j_1j_2}\right]\right\},
\end{aligned}
$$

and $s_{2,n}(\overline{\sigma}) = s_{i_1,j_1}(\overline{\sigma}) - s_{i_1,j_2}(\overline{\sigma}) - s_{i_2,j_1}(\overline{\sigma}) + s_{i_2,j_2}(\overline{\sigma})$. It follows then that $S^{\dagger}_{2,n\tau}$ can be written as

$$
\begin{aligned}
S^{\dagger}_{2,n\tau} &= \left[6!\binom{n}{6}\right]^{-1}\sum_{\overline{\sigma}\in\mathcal{N}_{\overline{m}_n}} s_{2,n\tau}(\overline{\sigma}) \\
&= \left[6!\binom{n}{6}\right]^{-1}\sum_{\overline{\sigma}\in\mathcal{N}_{\overline{m}_n}} \left\{s_{i_1j_1}(\overline{\sigma}) - s_{i_1j_2}(\overline{\sigma}) - s_{i_2j_1}(\overline{\sigma}) + s_{i_2j_2}(\overline{\sigma})\right\}.
\end{aligned}
$$

Step 1. Hájek Projection

The rest of the proof makes use of the following index notation for dyads. Given the total number of ordered dyads $\mathbf{\overline{n}} = n(n-1)$, let the boldface indeces $\boldsymbol{\pi} = \mathbf{1}, \mathbf{2}, \cdots$ index the $\mathbf{\overline{n}}$ ordered dyads in the sample. In an abuse of notation, also let $\boldsymbol{\pi}$ denote the set $\{i_1, j_1\}$, where i_1 and j_1 are the indices that comprise dyad $\boldsymbol{\pi}$. In particular, $\boldsymbol{\pi}(1) = i_1$ and $\boldsymbol{\pi}(2) = j_1$, when $\boldsymbol{\pi} = \{i_1, j_1\}$.

With this notation at hand, $S^{\dagger}_{2,n\tau}$ can be expressed as

$$S^{\dagger}_{2,n\tau} = \left[6!\binom{n}{6}\right]^{-1} \sum_{\boldsymbol{\pi}_1=\mathbf{1}}^{\mathbf{\overline{n}}} \sum_{\boldsymbol{\pi}_2 \neq \boldsymbol{\pi}_1} \sum_{\boldsymbol{\pi}_3 \neq \boldsymbol{\pi}_1} \left\{ s_{\boldsymbol{\pi}_1}(\overline{\sigma}) - s_{\boldsymbol{\pi}_1(1)\boldsymbol{\pi}_2(2)}(\overline{\sigma}) - s_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_1(2)}(\overline{\sigma}) + s_{\boldsymbol{\pi}_2}(\overline{\sigma}) \right\}$$

where $\overline{\sigma} = \overline{\sigma}\left\{\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \boldsymbol{\pi}_3\right\}$.

Let

$$\begin{aligned}
p_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_3}(\overline{\sigma}) &= \frac{1}{h^L}\left(\frac{\varphi_{\boldsymbol{\pi}_1,\tau}}{f_{vx,\boldsymbol{\pi}_1}}\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2} + \frac{\varphi_{\boldsymbol{\pi}_3,\tau}}{f_{vx,\boldsymbol{\pi}_3}}\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}\right) K_{x,h}\left(X_{\boldsymbol{\pi}_3} - X_{\boldsymbol{\pi}_1}\right) \\
&\quad - E\left[\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2} D^*_{\boldsymbol{\pi}_1,\tau} \mid \Omega_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\right] - E\left[\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2} D^*_{\boldsymbol{\pi}_3,\tau} \mid \Omega_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}\right] \\
p_{\boldsymbol{\pi}_2,\boldsymbol{\pi}_3}(\overline{\sigma}) &= \frac{1}{h^L}\left(\frac{\varphi_{\boldsymbol{\pi}_2,\tau}}{f_{vx,\boldsymbol{\pi}_2}}\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2} K_{x,h}\left(X_{\boldsymbol{\pi}_3} - X_{\boldsymbol{\pi}_2}\right) + \frac{\varphi_{\boldsymbol{\pi}_2,\tau}}{f_{vx,\boldsymbol{\pi}_2}}\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2} K_{x,h}\left(X_{\boldsymbol{\pi}_1} - X_{\boldsymbol{\pi}_2}\right)\right) \\
&\quad - E\left[\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2} D^*_{\boldsymbol{\pi}_2,\tau} \mid \Omega_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\right] - E\left[\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2} D^*_{\boldsymbol{\pi}_2,\tau} \mid \Omega_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}\right] \\
p_{\boldsymbol{\pi}_1(1)\boldsymbol{\pi}_2(2),\boldsymbol{\pi}_3}(\overline{\sigma}) &= \frac{1}{h^L}\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\left\{\left(\frac{\varphi_{\boldsymbol{\pi}_1(1)\boldsymbol{\pi}_2(2),\tau}}{f_{vx,\boldsymbol{\pi}_1(1)\boldsymbol{\pi}_2(2)}}\right) K_{x,h}\left(X_{\boldsymbol{\pi}_3} - X_{\boldsymbol{\pi}_1(1)\boldsymbol{\pi}_2(2)}\right) - E\left[D^*_{\boldsymbol{\pi}_1(1)\boldsymbol{\pi}_2(2),\tau} \mid \Omega_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\right]\right\} \\
&\quad + \frac{1}{h^L}\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}\left\{\left(\frac{\varphi_{\boldsymbol{\pi}_3(1)\boldsymbol{\pi}_2(2),\tau}}{f_{vx,\boldsymbol{\pi}_3(1)\boldsymbol{\pi}_2(2)}} K_{x,h}\left(X_{\boldsymbol{\pi}_1} - X_{\boldsymbol{\pi}_3(1)\boldsymbol{\pi}_2(2)}\right)\right) - E\left[D^*_{\boldsymbol{\pi}_3(1)\boldsymbol{\pi}_2(2),\tau} \mid \Omega_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}\right]\right\} \\
p_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_1(2),\boldsymbol{\pi}_3}(\overline{\sigma}) &= \frac{1}{h^L}\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\left\{\left(\frac{\varphi_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_1(2),\tau}}{f_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_1(2),\tau}} K_{x,h}\left(X_{\boldsymbol{\pi}_3} - X_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_1(2)}\right)\right) - E\left[D^*_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_1(2),\tau} \mid \Omega_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\right]\right\} \\
&\quad + \frac{1}{h^L}\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}\left\{\left(\frac{\varphi_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_3(2),\tau}}{f_{vx,\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_3(2)}} K_{x,h}\left(X_{\boldsymbol{\pi}_1} - X_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_3(2)}\right)\right) - E\left[D^*_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_3(2),\tau} \mid \Omega_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}\right]\right\}
\end{aligned}$$

where $K_{x,h}\left(X_{\boldsymbol{\pi}_3} - X_{\boldsymbol{\pi}_1}\right)$ denotes $K_{x,h}\left(X_{\boldsymbol{\pi}_3(1)} - X_{\boldsymbol{\pi}_1(1)}, X_{\boldsymbol{\pi}_3(2)} - X_{\boldsymbol{\pi}_1(2)}\right)$, $\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}$ denotes $\tilde{W}_{\boldsymbol{\pi}_1\{i_1 i_2\},\boldsymbol{\pi}_2\{j_1 j_2\}}$, and

$$\begin{aligned}
\chi_{\boldsymbol{\pi}_1} &= E\left[\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2} \mid X_{\boldsymbol{\pi}_1}\right] \\
\chi_{\boldsymbol{\pi}_1} &= \sum_{\boldsymbol{\pi}_2 \neq \boldsymbol{\pi}_1,\boldsymbol{\pi}_3} \chi_{\boldsymbol{\pi}_1}.
\end{aligned}$$

Using the symmetry of the kernel,it follows that $S^{\dagger}_{2,n\tau}$ can be written as

$$\left[6!\binom{n}{6}\right]^{-1} \sum_{\boldsymbol{\pi}_1=\mathbf{1}}^{\mathbf{\overline{n}}} \sum_{\boldsymbol{\pi}_3=\boldsymbol{\pi}_1+1} \sum_{\boldsymbol{\pi}_2 \neq \boldsymbol{\pi}_1,\boldsymbol{\pi}_3} \left\{ p_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_3}(\overline{\sigma}) - p_{\boldsymbol{\pi}_1(1)\boldsymbol{\pi}_2(2),\boldsymbol{\pi}_3}(\overline{\sigma}) - p_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_1(2),\boldsymbol{\pi}_3}(\overline{\sigma}) + p_{\boldsymbol{\pi}_2,\boldsymbol{\pi}_3}(\overline{\sigma}) \right\}$$

To compute the Hájek projection of the sum above into an an arbitraty function of $\zeta_{\boldsymbol{\pi}_1}$, consider first $E\left[p_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_3}\left(\overline{\sigma}\right)\mid\zeta_{\boldsymbol{\pi}_1}\right]$. To that end, the following results will be useful.

$$\begin{aligned}
E\left[E\left[\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}D^*_{\boldsymbol{\pi}_1,\tau}\mid\omega_{\boldsymbol{\pi}_1}\right]\mid\zeta_{\boldsymbol{\pi}_1}\right] &= E\left[D^*_{\boldsymbol{\pi}_1,\tau}\mid\omega_{\boldsymbol{\pi}_1}\right]E\left[\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\mid X_{\boldsymbol{\pi}_1}\right]=E\left[D^*_{\boldsymbol{\pi}_1,\tau}\chi_{\boldsymbol{\pi}_1}\mid\omega_{\boldsymbol{\pi}_1}\right]\\
E\left[E\left[\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}D^*_{\boldsymbol{\pi}_3,\tau}\mid\omega_{\boldsymbol{\pi}_3}\right]\mid\zeta_{\boldsymbol{\pi}_1}\right] &= E\left[E\left[D^*_{\boldsymbol{\pi}_3,\tau}\mid\omega_{\boldsymbol{\pi}_3}\right]E\left[\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}\mid X_{\boldsymbol{\pi}_3}\right]\right]=E\left[D^*_{\boldsymbol{\pi}_3,\tau}\chi_{\boldsymbol{\pi}_3}\right].
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&E\left[\left(\frac{\varphi_{\boldsymbol{\pi}_1,\tau}}{f_{vx,\boldsymbol{\pi}_1}}\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}+\frac{\varphi_{\boldsymbol{\pi}_3,\tau}}{f_{vx,\boldsymbol{\pi}_3}}\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}\right)\frac{1}{h^L}K_{x,h}\left(X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_1}\right)\mid\zeta_{\boldsymbol{\pi}_1}\right]\\
=&E\left[\left\{\frac{\varphi_{\boldsymbol{\pi}_1,\tau}}{f_{vx,\boldsymbol{\pi}_1}}E\left[\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\mid X_{\boldsymbol{\pi}_1}\right]+E\left[\frac{\varphi_{\boldsymbol{\pi}_3,\tau}}{f_{vx,\boldsymbol{\pi}_3}}\mid X_{\boldsymbol{\pi}_3}\right]E\left[\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}\mid X_{\boldsymbol{\pi}_3}\right]\right\}\frac{1}{h^L}K_{x,h}\left(X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_1}\right)\mid\zeta_{\boldsymbol{\pi}_1}\right]\\
=&\int\left\{\frac{\varphi_{\boldsymbol{\pi}_1,\tau}}{f_{vx,\boldsymbol{\pi}_1}}\chi_{\boldsymbol{\pi}_1}+E\left[\frac{\varphi_{\boldsymbol{\pi}_3,\tau}}{f_{vx,\boldsymbol{\pi}_3}}\chi_{\boldsymbol{\pi}_3}\mid X_{\boldsymbol{\pi}_3}\right]\right\}\frac{1}{h^L}K_{x,h}\left(X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_1}\right)f_x(X_{\boldsymbol{\pi}_3})dX_{\boldsymbol{\pi}_3}
\end{aligned}$$

where the second equality follows from a Law of Iterated Expectations and Assumption 3.1.1.

Let

$$\Xi\left(X_{\boldsymbol{\pi}_3}\right) = E\left[D^*_{\boldsymbol{\pi}_3,\tau}\chi_{\boldsymbol{\pi}_3}\mid X_{\boldsymbol{\pi}_3}\right],$$

and consider

$$\begin{aligned}
&\int\left\{\frac{\varphi_{\boldsymbol{\pi}_1,\tau}}{f_{vx,\boldsymbol{\pi}_1}}\chi_{\boldsymbol{\pi}_1}f_x(X_{\boldsymbol{\pi}_3})+\Xi\left(X_{\boldsymbol{\pi}_3}\right)\right\}\frac{1}{h^L}K_{x,h}\left(X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_1}\right)dX_{\boldsymbol{\pi}_3}-\left\{\frac{\varphi_{\boldsymbol{\pi}_1,\tau}}{f_{vx,\boldsymbol{\pi}_1}}\chi_{\boldsymbol{\pi}_1}f_x(X_{\boldsymbol{\pi}_1})+\Xi\left(X_{\boldsymbol{\pi}_1}\right)\right\}\\
=&\int\left\{\frac{\varphi_{\boldsymbol{\pi}_1,\tau}}{f_{vx,\boldsymbol{\pi}_1}}\chi_{\boldsymbol{\pi}_1}f_x(X_{\boldsymbol{\pi}_1}+h\boldsymbol{\nu})+\Xi\left(X_{\boldsymbol{\pi}_1}+h\boldsymbol{\nu}\right)\right\}K_{x,h}\left(\boldsymbol{\nu}\right)d\boldsymbol{\nu}-\left\{\frac{\varphi_{\boldsymbol{\pi}_1,\tau}}{f_{vx,\boldsymbol{\pi}_1}}\chi_{\boldsymbol{\pi}_1}f_x(X_{\boldsymbol{\pi}_1})+\Xi\left(X_{\boldsymbol{\pi}_1}\right)\right\}\\
=&\int\left\{\frac{\varphi_{\boldsymbol{\pi}_1,\tau}}{f_{vx,\boldsymbol{\pi}_1}}\chi_{\boldsymbol{\pi}_1}\left(f_x(X_{\boldsymbol{\pi}_1}+h\boldsymbol{\nu})-f_x(X_{\boldsymbol{\pi}_1})\right)\right\}+\left\{\Xi\left(X_{\boldsymbol{\pi}_1}+h\boldsymbol{\nu}\right)-\Xi\left(X_{\boldsymbol{\pi}_1}\right)\right\}K_x\left(\boldsymbol{\nu}\right)d\boldsymbol{\nu}\\
=&\,o(h^{\overline{M}})
\end{aligned}$$

where the first equality follows from a change of variable $\boldsymbol{\nu}=h^{-1}\left(X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_1}\right)$ with Jacobian h^L. The last equality follows Assumptions 4.1.1, 4.1.3, and 4.1.5 which guarantee that $f_x(X_{\boldsymbol{\pi}_1})$ and $\Xi\left(X_{\boldsymbol{\pi}_1}\right)$ are continuous and $\overline{M}$-times differentiable with respect to all of its arguments, and K_x is a bias-reducing kernel of order $2\overline{M}$. Observe that

$$\frac{\varphi_{\boldsymbol{\pi}_1,\tau}}{f_{vx,\boldsymbol{\pi}_1}}\chi_{\boldsymbol{\pi}_1}f_x(X_{\boldsymbol{\pi}_1}) = 0$$

holds for any $X_{\boldsymbol{\pi}_1}$ within a τ distance of the boundary $\mathbb{S}_x$, and having $h/\tau\to 0$ ensures that the change of variable $\boldsymbol{\nu}=h^{-1}\left(X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_1}\right)$ is not affected by boundary effects.

The previous results, and Assumption 4.1.5, yield

$$E\left[p_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_3}\left(\overline{\sigma}\right)\mid\zeta_{\boldsymbol{\pi}_1}\right]=D^*_{\boldsymbol{\pi}_1,\tau}\chi_{\boldsymbol{\pi}_1}+E\left[D^*_{\boldsymbol{\pi}_1,\tau}\chi_{\boldsymbol{\pi}_1}\mid X_{\boldsymbol{\pi}_1}\right]-E\left[D^*_{\boldsymbol{\pi}_1,\tau}\chi_{\boldsymbol{\pi}_1}\mid\omega_{\boldsymbol{\pi}_1}\right]-E\left[D^*_{\boldsymbol{\pi}_1,\tau}\chi_{\boldsymbol{\pi}_1}\right]+o(1).$$

Notice that for $\boldsymbol{\pi}_s \in \{(\boldsymbol{\pi}_1(1), \boldsymbol{\pi}_2(2)), (\boldsymbol{\pi}_2(1), \boldsymbol{\pi}_1(2)), \boldsymbol{\pi}_2\}$,

$$
\begin{aligned}
& E\left[\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\left\{\frac{1}{h^L}\frac{\varphi_{\boldsymbol{\pi}_s,\tau}}{f_{vx,\boldsymbol{\pi}_s}}K_{x,h}\left(X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_s}\right)-E\left[D^*_{\boldsymbol{\pi}_s,\tau}\mid\omega_{\boldsymbol{\pi}_s}\right]\right\}\mid\zeta_{\boldsymbol{\pi}_1}\right] \\
= \; & E\left[\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\left\{E\left[\frac{1}{h^L}\frac{\varphi_{\boldsymbol{\pi}_s,\tau}}{f_{vx,\boldsymbol{\pi}_s}}K_{x,h}\left(X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_s}\right)\mid\Omega_\sigma,\zeta_{\boldsymbol{\pi}_1}\right]-E\left[D^*_{\boldsymbol{\pi}_s,\tau}\mid\omega_{\boldsymbol{\pi}_s}\right]\right\}\mid\zeta_{\boldsymbol{\pi}_1}\right] \\
= \; & O\left(h^{\overline{M}}\right)
\end{aligned}
$$

since the expectation

$$
\begin{aligned}
E\left[\frac{1}{h^L}\frac{\varphi_{\boldsymbol{\pi}_s,\tau}}{f_{vx,\boldsymbol{\pi}_s}}K_{x,h}\left(X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_s}\right)\mid\Omega_\sigma,\zeta_{\boldsymbol{\pi}_1}\right] &= \int\frac{1}{h^L}E\left[\frac{\varphi_{\boldsymbol{\pi}_s,\tau}}{f_{vx,\boldsymbol{\pi}_s}}\mid\omega_{\boldsymbol{\pi}_s}\right]K_{x,h}\left(X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_s}\right)f_x\left(X_{\boldsymbol{\pi}_3}\right)dX_{\boldsymbol{\pi}_3} \\
&= E\left[D^*_{\boldsymbol{\pi}_s,\tau}\mid\omega_{\boldsymbol{\pi}_s}\right]+O\left(h^{\overline{M}}\right),
\end{aligned}
$$

where the second equality follows from Assumptions 3.1.1, 3.1.2, and properties of the bias-reducing kernel, Assumption 4.1.5.

Similarly, for a given $\boldsymbol{\pi}_s \in \{(\boldsymbol{\pi}_3(1), \boldsymbol{\pi}_2(2)), (\boldsymbol{\pi}_2(1), \boldsymbol{\pi}_3(2)), \boldsymbol{\pi}_2\}$, it follows from Assumptions 3.1.1, 3.1.2, 4.1.3, and 4.1.5, that

$$
\begin{aligned}
& E\left[\frac{1}{h^L}\left(\frac{\varphi_{\boldsymbol{\pi}_s,\tau}}{f_{vx,\boldsymbol{\pi}_s}}\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}K_{x,h}\left(X_{\boldsymbol{\pi}_1}-X_{\boldsymbol{\pi}_s}\right)\right)\mid\zeta_{\boldsymbol{\pi}_1}\right]-\Xi\left[X_{\boldsymbol{\pi}_1}\right] \\
= \; & E\left[\frac{1}{h^L}E\left[\left(\frac{\varphi_{\boldsymbol{\pi}_s,\tau}}{f_{vx,\boldsymbol{\pi}_s}}\chi_{\boldsymbol{\pi}_s}\right)\mid X_{\boldsymbol{\pi}_s}\right]K_{x,h}\left(X_{\boldsymbol{\pi}_1}-X_{\boldsymbol{\pi}_s}\right)\mid\zeta_{\boldsymbol{\pi}_1}\right]-\Xi\left[X_{\boldsymbol{\pi}_1}\right] \\
= \; & \int\left\{\Xi\left[X_{\boldsymbol{\pi}_1}+h\boldsymbol{\nu}\right]-\Xi\left[X_{\boldsymbol{\pi}_1}\right]\right\}K_x\left(\boldsymbol{\nu}\right)d\boldsymbol{\nu} \\
= \; & O\left(h^{\overline{M}}\right).
\end{aligned}
$$

Using the previous results it follows that

$$
E\left[p_{\boldsymbol{\pi}_s,\boldsymbol{\pi}_3}\left(\overline{\sigma}\right)\mid\zeta_{\boldsymbol{\pi}_1}\right] = E\left[D^*_{\boldsymbol{\pi}_1,\tau}\chi_{\boldsymbol{\pi}_1}\mid X_{\boldsymbol{\pi}_1}\right]-E\left[D^*_{\boldsymbol{\pi}_1,\tau}\chi_{\boldsymbol{\pi}_1}\right],
$$

and thus,

$$
\begin{aligned}
& E\left[p_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_3}\left(\overline{\sigma}\right)-p_{\boldsymbol{\pi}_1(1)\boldsymbol{\pi}_2(2),\boldsymbol{\pi}_3}\left(\overline{\sigma}\right)-p_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_1(2),\boldsymbol{\pi}_3}\left(\overline{\sigma}\right)+p_{\boldsymbol{\pi}_2,\boldsymbol{\pi}_3}\left(\overline{\sigma}\right)\mid\zeta_{\boldsymbol{\pi}_1}\right] \\
= \; & \left\{D^*_{\boldsymbol{\pi}_1}-E\left[D^*_{\boldsymbol{\pi}_1}\mid\omega_{\boldsymbol{\pi}_1}\right]\right\}I_{\tau,\boldsymbol{\pi}_1}\chi_{\boldsymbol{\pi}_1}+o(1)
\end{aligned}
$$

It follows then that the Hájek projection is given by

$$
V^*_{2,n\tau} = \frac{1}{n(n-1)}\sum_{i_1=1}^{n}\sum_{j_1\neq i_1}\xi_{i_1j_1,\tau}+o(1)
$$

with

$$\xi_{i_1 j_1,\tau} = \left\{ D^*_{i_1 j_1} - E\left[D^*_{i_1 j_1} \mid \omega_{i_1 j_1} \right] \right\} I_{\tau, i_1 j_1} \overline{\chi}_{i_1 j_1}$$

$$\overline{\chi}_{i_1 j_1} = \left\{ \frac{1}{(n-2)(n-3)} \sum_{i_2 \neq i_1, j_1} \sum_{j_2 \neq i_1, j_1, i_2} E\left[\tilde{W}_{\sigma\{i_1, i_2; j_1, j_2\}} \mid X_{i_1}, X_{j_1} \right] \right\}.$$

If follows from a Law of Iterated Expectations that

$$E\left[V^*_{2,n\tau} \right] = E\left[\xi_{i_1 j_1,\tau} \right] = 0.$$

Step 2. Variance of Hájek Projection

As in the proof of Lemma B.4, the variance of $V^*_{1,n\tau}$ is given by

$$\begin{aligned} Var\left(V^*_{1,n\tau} \right) &= \left\{ \frac{1}{n(n-1)} \right\}^2 \left\{ \sum_{i_1=1}^{n} \sum_{j_1 \neq i_1} E\left[\xi_{i_1 j_1,\tau} \xi'_{i'_1 j'_1,\tau} \right] \right\} \\ &= \left\{ \frac{1}{n(n-1)} \right\}^2 \left\{ \sum_{i_1=1}^{n} \sum_{j_1 \neq i_1} \Lambda^*_{i_1, j_1} \right\} \end{aligned}$$

where

$$\Lambda^*_{i_1, j_1} = E\left[\left\{ E\left[D^*_{i_1 j_1} D^*_{i_1 j_1} \mid \omega_{i_1 j_1} \right] - E\left[D^*_{i_1 j_1} \mid \omega_{i_1 j_1} \right]^2 \right\} I^2_{\tau, i_1 j_1} \overline{\chi}_{i_1 j_1} \overline{\chi}'_{i_1 j_1} \right].$$

Define

$$\Upsilon_n = n(n-1) Var\left(V^*_{1,n\tau} \right) = \frac{1}{n(n-1)} \left\{ \sum_{i_1=1}^{n} \sum_{j_1 \neq i_1} \Lambda^*_{i_1, j_1} \right\}.$$

Step 3. Variance of $S_{2,n\tau}$

Given two different 6-tuples $\overline{\sigma}\{i_1, i_2; j_1, j_2; l_1, l_2\}$ and $\overline{\sigma}'\{i'_1, i'_2; j'_1, j'_2; l'_1, l'_2\}$, let

$$\Delta_{c,n} = Cov\left(s_{2,n}\left(\sigma\{i_1, i_2; j_1, j_2; l_1, l_2\} \right), s_{2,n}\left(\sigma'\{i'_1, i'_2; j'_1, j'_2; l'_1, l'_2\} \right) \right)$$

denote the covariance between $s_{2,n}(\overline{\sigma})$ and $s_{2,n}(\overline{\sigma}')$ when $\overline{\sigma}$ and $\overline{\sigma}'$ have $c = 0, 1, 2, 3, 4, 5, 6$ indices in common.

It follows from conditionally independence formation of links, implied by Assumption 3.1.2, and the conditional mean zero, $E\left[s_{2,n}\left(\sigma\{i_1, i_2; j_1, j_2; l_1, l_2\} \right) \mid \Omega_\sigma \right] = 0$, that $\Delta_{0,n} = \Delta_{1,n} = 0$.

Consider

$$\begin{aligned}
\Delta_{2,n} &= E\left[s_{2,n}(\overline{\sigma}\{i_1,i_2;j_1,j_2;l_1,l_2\})s_{2,n}(\overline{\sigma'}\{i_1,i_2';j_1,j_2';l_1,l_2'\})'\right] \\
&= E\left[s_{i_1j_1}\left(\overline{\sigma}\right)s_{i_1j_1}\left(\overline{\sigma'}\right)'\right]+o(1) \\
&= E\left[\left\{E\left[\tilde{D}^*_{i_1j_1,\tau}\tilde{D}^*_{i_1j_1,\tau}\mid\omega_{i_1j_1}\right]-E\left[\tilde{D}^*_{i_1j_1,\tau}\mid\omega_{i_1j_1}\right]^2\right\}I^2_{\tau,i_1j_1}\tilde{W}_\sigma\tilde{W}_{\sigma'}\right]+o(1).
\end{aligned}$$

Therefore, the variance of $Var(S^\dagger_{2,n\tau})$ can be expressed as

$$\begin{aligned}
&\left(\frac{1}{m_n}\right)^2\sum_{\overline{\sigma}}\sum_{\overline{\sigma}'}E\left[\left(s_{2,n}(\overline{\sigma})s_{2,n}(\overline{\sigma'})'\right)\right] \\
&+\left(4!\binom{n}{4}\right)^{-2}\sum_{i_1=1}^n\sum_{j_1\neq i_1}\left\{\sum_{k_1\neq i_1,j_1}\sum_{k_2\neq i_1,j_1,k_1}\sum_{l_1\neq i_1,j_1}\sum_{l_2\neq i_1,j_1,l_1}\Delta_{2,n}\right\} \\
&+O\left(\frac{1}{n^3}\right)\Delta_{3,n}+O\left(\frac{1}{n^4}\right)\Delta_{4,n}+O\left(\frac{1}{n^5}\right)\Delta_{5,n}+O\left(\frac{1}{n^6}\right)\Delta_{6,n}
\end{aligned}$$

Notice that the term inside the brackets scaled by $((n-2)(n-3))^{-2}$ can be written as

$$\begin{aligned}
&\left(\frac{1}{(n-2)(n-3)}\right)^2\sum_{k_1\neq i_1,j_1}\sum_{k_2\neq i_1,j_1,k_1}\sum_{l_1\neq i_1,j_1}\sum_{l_2\neq i_1,j_1,l_1}\Delta_{2,n} \\
=\;& E\left[\left\{E\left[D^*_{i_1j_1}D^*_{i_1j_1}\mid\omega_{i_1j_1}\right]-E\left[D^*_{i_1j_1}\mid\omega_{i_1j_1}\right]^2\right\}I^2_{\tau,i_1j_1}\chi_{i_1j_1}\chi'_{i_1j_1}\right] \\
=\;& \Lambda^*_{i_1,j_1}.
\end{aligned}$$

As a result,

$$Var\left[S^\dagger_{2,n\tau}\right] = \left(\frac{1}{n(n-1)}\right)^2\left\{\sum_{i_1=1}^n\sum_{j_1\neq i_1}\Lambda^*_{i_1,j_1}\right\}+o(1),$$

and $Var\left[V^*_{2,n\tau}\right]-Var\left[S^\dagger_{2,n\tau}\right]=o_p(1)$.

The asymptotic equivalence results follows from similar arguments as in the proof of Lemma B.4. The proof is complete.

□

Lemma B.6. *Under the same Assumptions of Theorem 4.2, it follows that the Hájek projection of*

$$\begin{aligned}
S^\dagger_{3,n\tau} &= S_{3,n\tau}-E\left[\tilde{W}_\sigma\tilde{D}^*_{\sigma,\tau}\mid\Omega_\sigma\right] \\
S_{3,n\tau} &= \frac{1}{m_n}\sum_{\sigma\in\mathcal{N}_{m_n}}\tilde{W}_\sigma\left\{\left(\frac{D^*_{i_1j_1,\tau}\widehat{f}_{vx,i_1j_1}}{f_{vx,i_1j_1}}-\frac{D^*_{i_1j_2,\tau}\widehat{f}_{vx,i_1j_2}}{f_{vx,i_1j_2}}\right)-\left(\frac{D^*_{i_2j_1,\tau}\widehat{f}_{vx,i_2j_1}}{f_{vx,i_2j_1}}-\frac{D^*_{i_2j_2,\tau}\widehat{f}_{vx,i_2j_2}}{f_{vx,i_2j_2}}\right)\right\}
\end{aligned}$$

into an arbitrary function $\zeta_{i_1 j_1} = (X_{i_1}, X_{j_1}, A_{i_1}, A_{j_1}, v_{i_1 j_1}, U_{i_1 j_1})$ is given by

$$V^*_{3,n\tau} \;=\; E\left[S^\dagger_{3,n\tau} \mid \zeta_{i_1 j_1}\right] \;=\; \frac{1}{n(n-1)} \sum_{i_1=1}^{n} \sum_{j_1 \neq i_1} \xi_{i_1 j_1,\tau}$$

and

$$n(n-1)\Upsilon_n^{-1/2} E\left[\left(S^\dagger_{3,n\tau} - V^*_{3,n\tau}\right)^2\right] \Upsilon_n^{-1/2} = o(1),$$

where $\Upsilon_n = n(n-1)Var(V^*_{3,n\tau})$.

Proof. Consider a fixed 6-tuple $\{i_1, i_2; j_1, j_2; k_1, k_2\}$, and define

$$\begin{aligned}
s_{i_1,j_1}(\overline{\sigma}) &= \tilde{W}_{i_1 i_2, j_1 j_2}\left\{\frac{1}{h^{L+1}}\frac{D^*_{i_1 j_1,\tau}}{f_{vx,i_1 j_1}} K_{vx,h}\left(v_{k_1 k_2} - v_{i_1 j_1}, X_{k_1} - X_{i_1}, X_{k_2} - X_{j_1}\right) - E\left[D^*_{i_1 j_1,\tau} \mid \Omega_{i_1 i_2, j_1 j_2}\right]\right\} \\
s_{i_1,j_2}(\overline{\sigma}) &= \tilde{W}_{i_1 i_2, j_1 j_2}\left\{\frac{1}{h^{L+1}}\frac{D^*_{i_1 j_2,\tau}}{f_{vx,i_1 j_2}} K_{vx,h}\left(v_{k_1 k_2} - v_{i_1 j_2}, X_{k_1} - X_{i_1}, X_{k_2} - X_{j_2}\right) - E\left[D^*_{i_1 j_2,\tau} \mid \Omega_{i_1 i_2, j_1 j_2}\right]\right\} \\
s_{i_2,j_1}(\overline{\sigma}) &= \tilde{W}_{i_1 i_2, j_1 j_2}\left\{\frac{1}{h^{L+1}}\frac{D_{i_2 j_1,\tau^*}}{f_{vx,i_2 j_1}} K_{vx,h}\left(v_{k_1 k_2} - v_{i_2 j_1}, X_{k_1} - X_{i_2}, X_{k_2} - X_{j_1}\right) - E\left[D^*_{i_2 j_1,\tau} \mid \Omega_{i_1 i_2, j_1 j_2}\right]\right\} \\
s_{i_2,j_2}\overline{\sigma}) &= \tilde{W}_{i_1 i_2, j_1 j_2}\left\{\frac{1}{h^{L+1}}\frac{D^*_{i_2 j_2,\tau}}{f_{vx,i_2 j_2}} K_{vx,h}\left(v_{k_1 k_2} - v_{i_2 j_2}, X_{k_1} - X_{i_2}, X_{k_2} - X_{j_2}\right) - E\left[D^*_{i_2 j_2,\tau} \mid \Omega_{i_1 i_2, j_1 j_2}\right]\right\},
\end{aligned}$$

and $s_{3,n}(\overline{\sigma}) = s_{i_1,j_1}(\overline{\sigma}) - s_{i_1,j_2}(\overline{\sigma}) - s_{i_2,j_1}(\overline{\sigma}) + s_{i_2,j_2}(\overline{\sigma})$. It follows then that $S^\dagger_{3,n\tau}$ can be written as

$$\begin{aligned}
S^\dagger_{3,n\tau} &= \left[6!\binom{n}{6}\right]^{-1} \sum_{\overline{\sigma} \in \mathcal{N}_{\overline{m}n}} s_{2,n\tau}(\overline{\sigma}) \\
&= \left[6!\binom{n}{6}\right]^{-1} \sum_{\overline{\sigma} \in \mathcal{N}_{\overline{m}n}} \left\{s_{i_1 j_1}(\overline{\sigma}) - s_{i_1 j_2}(\overline{\sigma}) - s_{i_2 j_1}(\overline{\sigma}) + s_{i_2 j_2}(\overline{\sigma})\right\}.
\end{aligned}$$

Step 1. Hájek Projection

The rest of the proof makes use of the following index notation for dyads. Given the total number of ordered dyads $\overline{\mathbf{n}} = n(n-1)$, let the boldface indeces $\boldsymbol{\pi} = \mathbf{1}, \mathbf{2}, \cdots$ index the $\overline{\mathbf{n}}$ ordered dyads in the sample. In an abuse of notation, also let $\boldsymbol{\pi}$ denote the set $\{i_1, j_1\}$, where i_1 and j_1 are the indices that comprise dyad $\boldsymbol{\pi}$. In particular, $\boldsymbol{\pi}(1) = i_1$ and $\boldsymbol{\pi}(2) = j_1$, when $\boldsymbol{\pi} = \{i_1, j_1\}$.

With this notation at hand, $S^\dagger_{3,n\tau}$ can be expressed as

$$S^\dagger_{3,n\tau} = \left[6!\binom{n}{6}\right]^{-1} \sum_{\boldsymbol{\pi}_1 = \mathbf{1}}^{\overline{\mathbf{n}}} \sum_{\boldsymbol{\pi}_2 \neq \boldsymbol{\pi}_1} \sum_{\boldsymbol{\pi}_3 \neq \boldsymbol{\pi}_1} \left\{s_{\boldsymbol{\pi}_1}(\overline{\sigma}) - s_{\boldsymbol{\pi}_1(1)\boldsymbol{\pi}_2(2)}(\overline{\sigma}) - s_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_1(2)}(\overline{\sigma}) + s_{\boldsymbol{\pi}_2}(\overline{\sigma})\right\}$$

where $\overline{\sigma} = \overline{\sigma}\{\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \boldsymbol{\pi}_3\}$.

Let

$$\begin{aligned}
p_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_3}(\overline{\sigma}) &= \frac{1}{h^{L+1}}\left(\frac{D^*_{\boldsymbol{\pi}_1,\tau}}{f_{vx,\boldsymbol{\pi}_1}}\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}+\frac{D^*_{\boldsymbol{\pi}_3,\tau}}{f_{vx,\boldsymbol{\pi}_3}}\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}\right)K_{vx,h}\left(v_{\boldsymbol{\pi}_3}-v_{\boldsymbol{\pi}_1},X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_1}\right)\\
&\quad -E\left[\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}D^*_{\boldsymbol{\pi}_1,\tau}\mid\Omega_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\right]-E\left[\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}D^*_{\boldsymbol{\pi}_3,\tau}\mid\Omega_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}\right]\\
p_{\boldsymbol{\pi}_2,\boldsymbol{\pi}_3}(\overline{\sigma}) &= \frac{1}{h^{L+1}}\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\left\{\frac{D^*_{\boldsymbol{\pi}_2,\tau}}{f_{vx,\boldsymbol{\pi}_2}}K_{vx,h}\left(v_{\boldsymbol{\pi}_3}-v_{\boldsymbol{\pi}_2},X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_2}\right)-E\left[D^*_{\boldsymbol{\pi}_2,\tau}\mid\Omega_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\right]\right\}\\
&\quad \frac{1}{h^{L+1}}\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}\left\{\frac{D^*_{\boldsymbol{\pi}_2,\tau}}{f_{vx,\boldsymbol{\pi}_2}}K_{vx,h}\left(v_{\boldsymbol{\pi}_1}-v_{\boldsymbol{\pi}_2},X_{\boldsymbol{\pi}_1}-X_{\boldsymbol{\pi}_2}\right)-E\left[D^*_{\boldsymbol{\pi}_2,\tau}\mid\Omega_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}\right]\right\}\\
p_{\boldsymbol{\pi}_1(1)\boldsymbol{\pi}_2(2),\boldsymbol{\pi}_3}(\overline{\sigma}) &= \frac{1}{h^{L+1}}\frac{D^*_{\boldsymbol{\pi}_1(1)\boldsymbol{\pi}_2(2),\tau}}{f_{vx,\boldsymbol{\pi}_1(1)\boldsymbol{\pi}_2(2)}}\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}K_{vx,h}\left(v_{\boldsymbol{\pi}_3}-v_{\boldsymbol{\pi}_1(1)\boldsymbol{\pi}_2(2)},X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_1(1)\boldsymbol{\pi}_2(2)}\right)\\
&\quad +\frac{1}{h^{L+1}}\frac{D^*_{\boldsymbol{\pi}_3(1)\boldsymbol{\pi}_2(2),\tau}}{f_{vx,\boldsymbol{\pi}_3(1)\boldsymbol{\pi}_2(2)}}\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}K_{vx,h}\left(v_{\boldsymbol{\pi}_1}-v_{\boldsymbol{\pi}_3(1)\boldsymbol{\pi}_2(2)},X_{\boldsymbol{\pi}_1}-X_{\boldsymbol{\pi}_3(1)\boldsymbol{\pi}_2(2)}\right)\\
&\quad -E\left[\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}D^*_{\boldsymbol{\pi}_1(1)\boldsymbol{\pi}_2(2),\tau}\mid\Omega_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\right]-E\left[\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}D^*_{\boldsymbol{\pi}_3(1)\boldsymbol{\pi}_2(2),\tau}\mid\Omega_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}\right]\\
p_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_1(2),\boldsymbol{\pi}_3}(\overline{\sigma}) &= \frac{1}{h^{L+1}}\frac{D^*_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_1(2),\tau}}{f_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_1(2),\tau}}\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}K_{vx,h}\left(v_{\boldsymbol{\pi}_3}-v_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_1(2)},X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_1(2)}\right)\\
&\quad +\frac{1}{h^{L+1}}\frac{D^*_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_3(2),\tau}}{f_{vx,\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_3(2)}}\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}K_{vx,h}\left(v_{\boldsymbol{\pi}_1}-v_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_3(2)},X_{\boldsymbol{\pi}_1}-X_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_3(2)}\right)\\
&\quad -E\left[\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}D^*_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_1(2),\tau}\mid\Omega_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\right]-E\left[\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}D^*_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_3(2),\tau}\mid\Omega_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}\right]
\end{aligned}$$

where $K_{vx,h}\left(v_{\boldsymbol{\pi}_3}-v_{\boldsymbol{\pi}_1},X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_1}\right)$ denotes $K_{vx,h}\left(v_{\boldsymbol{\pi}_3}-v_{\boldsymbol{\pi}_1},X_{\boldsymbol{\pi}_3(1)}-X_{\boldsymbol{\pi}_1(1)},X_{\boldsymbol{\pi}_3(2)}-X_{\boldsymbol{\pi}_1(2)}\right)$, $\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}$ denotes $\tilde{W}_{\boldsymbol{\pi}_1\{i_1i_2\},\boldsymbol{\pi}_2\{j_1j_2\}}$, and

$$\begin{aligned}
\chi_{\boldsymbol{\pi}_1} &= E\left[\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\mid X_{\boldsymbol{\pi}_1}\right]\\
\overline{\chi}_{\boldsymbol{\pi}_1} &= \sum_{\boldsymbol{\pi}_2\neq\boldsymbol{\pi}_1,\boldsymbol{\pi}_3}\chi_{\boldsymbol{\pi}_1}.
\end{aligned}$$

Using the symmetry of the kernel, it follows that $S^{\dagger}_{3,n\tau}$ can be written as

$$\left[6!\binom{n}{6}\right]^{-1}\sum_{\boldsymbol{\pi}_1=\mathbf{1}}^{\overline{\mathbf{n}}}\sum_{\boldsymbol{\pi}_3=\boldsymbol{\pi}_1+1}\sum_{\boldsymbol{\pi}_2\neq\boldsymbol{\pi}_1,\boldsymbol{\pi}_3}\left\{p_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_3}(\overline{\sigma})-p_{\boldsymbol{\pi}_1(1)\boldsymbol{\pi}_2(2),\boldsymbol{\pi}_3}(\overline{\sigma})-p_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_1(2),\boldsymbol{\pi}_3}(\overline{\sigma})+p_{\boldsymbol{\pi}_2,\boldsymbol{\pi}_3}(\overline{\sigma})\right\}$$

To compute the Hájek projection of the sum above into an an arbitraty function of $\zeta_{\boldsymbol{\pi}_1}$, consider first $E\left[p_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_3}(\overline{\sigma})\mid\zeta_{\boldsymbol{\pi}_1}\right]$. To that end, the following results will be useful.

$$\begin{aligned}
E\left[E\left[\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}D^*_{\boldsymbol{\pi}_1,\tau}\mid\omega_{\boldsymbol{\pi}_1}\right]\mid\zeta_{\boldsymbol{\pi}_1}\right] &= E\left[D^*_{\boldsymbol{\pi}_1,\tau}\mid\omega_{\boldsymbol{\pi}_1}\right]E\left[\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\mid X_{\boldsymbol{\pi}_1}\right]=E\left[D^*_{\boldsymbol{\pi}_1,\tau}\chi_{\boldsymbol{\pi}_1}\mid\omega_{\boldsymbol{\pi}_1}\right]\\
E\left[E\left[\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}D^*_{\boldsymbol{\pi}_3,\tau}\mid\omega_{\boldsymbol{\pi}_3}\right]\mid\zeta_{\boldsymbol{\pi}_1}\right] &= E\left[E\left[D^*_{\boldsymbol{\pi}_3,\tau}\mid\omega_{\boldsymbol{\pi}_3}\right]E\left[\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}\mid X_{\boldsymbol{\pi}_3}\right]\right]=E\left[D^*_{\boldsymbol{\pi}_3,\tau}\chi_{\boldsymbol{\pi}_3}\right].
\end{aligned}$$

Moreover,

$$
\begin{aligned}
& E\left[\left(\frac{D^*_{\boldsymbol{\pi}_1,\tau}}{f_{vx,\boldsymbol{\pi}_1}}\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}+\frac{D^*_{\boldsymbol{\pi}_3,\tau}}{f_{vx,\boldsymbol{\pi}_3}}\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}\right)\frac{1}{h^{L+1}}K_{vx,h}\left(v_{\boldsymbol{\pi}_3}-v_{\boldsymbol{\pi}_1},X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_1}\right)\mid\zeta_{\boldsymbol{\pi}_1}\right]\\
=& E\left[\left\{\frac{D^*_{\boldsymbol{\pi}_1,\tau}}{f_{vx,\boldsymbol{\pi}_1}}E\left[\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\mid X_{\boldsymbol{\pi}_1}\right]+E\left[\frac{D^*_{\boldsymbol{\pi}_3,\tau}}{f_{vx,\boldsymbol{\pi}_3}}\mid v_{\boldsymbol{\pi}_3},X_{\boldsymbol{\pi}_3}\right]E\left[\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}\mid X_{\boldsymbol{\pi}_3}\right]\right\}\right.\\
&\left.\quad\times\frac{1}{h^{L+1}}K_{vx,h}\left(v_{\boldsymbol{\pi}_3}-v_{\boldsymbol{\pi}_1},X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_1}\right)\mid\zeta_{\boldsymbol{\pi}_1}\right]\\
=& \int\left\{\frac{D^*_{\boldsymbol{\pi}_1,\tau}}{f_{vx,\boldsymbol{\pi}_1}}\chi_{\boldsymbol{\pi}_1}+E\left[\frac{D^*_{\boldsymbol{\pi}_3,\tau}}{f_{vx,\boldsymbol{\pi}_3}}\chi_{\boldsymbol{\pi}_3}\mid v_{\boldsymbol{\pi}_3},X_{\boldsymbol{\pi}_3}\right]\right\}\frac{1}{h^{L+1}}K_{vx,h}\left(v_{\boldsymbol{\pi}_3}-v_{\boldsymbol{\pi}_1},X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_1}\right)f_{vx}(v_{\boldsymbol{\pi}_3},X_{\boldsymbol{\pi}_3})dv_{\boldsymbol{\pi}_3}dX_{\boldsymbol{\pi}_3}
\end{aligned}
$$

where the second equality follows from a Law of Iterated Expectations and Assumption 3.1.1.

Let

$$
\Xi\left(v_{\boldsymbol{\pi}_3},X_{\boldsymbol{\pi}_3}\right) \quad=\quad E\left[D^*_{\boldsymbol{\pi}_3,\tau}\chi_{\boldsymbol{\pi}_3}\mid v_{\boldsymbol{\pi}_3},X_{\boldsymbol{\pi}_3}\right],
$$

and consider

$$
\begin{aligned}
& \int\left\{\frac{D^*_{\boldsymbol{\pi}_1,\tau}}{f_{vx,\boldsymbol{\pi}_1}}\chi_{\boldsymbol{\pi}_1}f_{vx}(v_{\boldsymbol{\pi}_3},X_{\boldsymbol{\pi}_3})+\Xi\left(v_{\boldsymbol{\pi}_3},X_{\boldsymbol{\pi}_3}\right)\right\}\frac{1}{h^{L+1}}K_{vx,h}\left(v_{\boldsymbol{\pi}_3}-v_{\boldsymbol{\pi}_1},X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_1}\right)dv_{\boldsymbol{\pi}_3}dX_{\boldsymbol{\pi}_3}\\
& -\left\{\frac{D^*_{\boldsymbol{\pi}_1,\tau}}{f_{vx,\boldsymbol{\pi}_1}}\chi_{\boldsymbol{\pi}_1}f_{vx}(v_{\boldsymbol{\pi}_1},X_{\boldsymbol{\pi}_1})+\Xi\left(v_{\boldsymbol{\pi}_1},X_{\boldsymbol{\pi}_1}\right)\right\}\\
= & \int\left\{\frac{D^*_{\boldsymbol{\pi}_1,\tau}}{f_{vx,\boldsymbol{\pi}_1}}\chi_{\boldsymbol{\pi}_1}f_{vx}(v_{\boldsymbol{\pi}_1}+h\boldsymbol{\nu_1}X_{\boldsymbol{\pi}_1}+h\boldsymbol{\nu_2})+\Xi\left(v_{\boldsymbol{\pi}_1}+h\boldsymbol{\nu_1},X_{\boldsymbol{\pi}_1}+h\boldsymbol{\nu_2}\right)\right\}K_{vx}\left(\boldsymbol{\nu}\right)d\boldsymbol{\nu}\\
& -\left\{\frac{D^*_{\boldsymbol{\pi}_1,\tau}}{f_{vx,\boldsymbol{\pi}_1}}\chi_{\boldsymbol{\pi}_1}f_{vx}(v_{\boldsymbol{\pi}_1},X_{\boldsymbol{\pi}_1})+\Xi\left(v_{\boldsymbol{\pi}_1},X_{\boldsymbol{\pi}_1}\right)\right\}\\
= & \int\left(\frac{D^*_{\boldsymbol{\pi}_1,\tau}}{f_{vx,\boldsymbol{\pi}_1}}\chi_{\boldsymbol{\pi}_1}\left\{f_{vx}(v_{\boldsymbol{\pi}_1}+h\boldsymbol{\nu_1}X_{\boldsymbol{\pi}_1}+h\boldsymbol{\nu_2})-f_{vx}(v_{\boldsymbol{\pi}_1},X_{\boldsymbol{\pi}_1})\right\}\right.\\
& \left.+\left\{\Xi\left(v_{\boldsymbol{\pi}_1}+h\boldsymbol{\nu_1},X_{\boldsymbol{\pi}_1}+h\boldsymbol{\nu_2}\right)-\Xi\left(v_{\boldsymbol{\pi}_1},X_{\boldsymbol{\pi}_1}\right)\right\}\right)K_{vx}\left(\boldsymbol{\nu}\right)d\boldsymbol{\nu}\\
= & o(h^{\overline{M}})
\end{aligned}
$$

where the first equality follows from a change of variable $\boldsymbol{\nu}=(\boldsymbol{\nu}_1,\boldsymbol{\nu}_2)$, with $\boldsymbol{\nu}_1=h^{-1}\left(v_{\boldsymbol{\pi}_3}-v_{\boldsymbol{\pi}_1}\right)$, and $\boldsymbol{\nu}_2=h^{-1}\left(X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_1}\right)$, with Jacobian h^L. The last equality follows Assumptions 4.1.1, 4.1.3, and 4.1.5 which guarantee that $f_{vx}(v_{\boldsymbol{\pi}_1},X_{\boldsymbol{\pi}_1})$ and $\Xi\left(v_{\boldsymbol{\pi}_1},X_{\boldsymbol{\pi}_1}\right)$ are continuous and $\overline{M}$-times differentiable with respect to all of its arguments, and K_{vx} is a bias-reducing kernel of order $2\overline{M}$. Observe that

$$
\frac{D^*_{\boldsymbol{\pi}_1,\tau}}{f_{vx,\boldsymbol{\pi}_1}}\chi_{\boldsymbol{\pi}_1}f_{vx}(v_{\boldsymbol{\pi}_1},X_{\boldsymbol{\pi}_1}) \quad=\quad 0
$$

holds for any $(v_{\boldsymbol{\pi}_1},X_{\boldsymbol{\pi}_1})$ within a τ distance of the boundary $\mathbb{S}_{vx}$, and having $h/\tau\to 0$ ensures that the change of variable $\boldsymbol{\nu}=(\boldsymbol{\nu}_1,\boldsymbol{\nu}_2)$, with $\boldsymbol{\nu}_1=h^{-1}\left(v_{\boldsymbol{\pi}_3}-v_{\boldsymbol{\pi}_1}\right)$, and $\boldsymbol{\nu}_2=h^{-1}\left(X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_1}\right)$, is not affected by boundary effects.

The previous results yield

$$
E\left[p_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_3}\left(\overline{\sigma}\right)\mid\zeta_{\boldsymbol{\pi}_1}\right]=D^*_{\boldsymbol{\pi}_1,\tau}\chi_{\boldsymbol{\pi}_1}+E\left[D^*_{\boldsymbol{\pi}_1,\tau}\chi_{\boldsymbol{\pi}_1}\mid X_{\boldsymbol{\pi}_1}\right]-E\left[D^*_{\boldsymbol{\pi}_1,\tau}\chi_{\boldsymbol{\pi}_1}\mid\omega_{\boldsymbol{\pi}_1}\right]-E\left[D^*_{\boldsymbol{\pi}_1,\tau}\chi_{\boldsymbol{\pi}_1}\right]+o(1).
$$

Notice that for $\boldsymbol{\pi}_s \in \{(\boldsymbol{\pi}_1(1), \boldsymbol{\pi}_2(2)), (\boldsymbol{\pi}_2(1), \boldsymbol{\pi}_1(2)), \boldsymbol{\pi}_2\}$,

$$
\begin{aligned}
& E\left[\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\left\{\frac{1}{h^{L+1}}\frac{D^*_{\boldsymbol{\pi}_s,\tau}}{f_{vx,\boldsymbol{\pi}_s}}K_{vx,h}\left(v_{\boldsymbol{\pi}_3}-v_{\boldsymbol{\pi}_s}, X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_s}\right)-E\left[D^*_{\boldsymbol{\pi}_s,\tau}\mid\omega_{\boldsymbol{\pi}_s}\right]\right\}\mid\zeta_{\boldsymbol{\pi}_1}\right] \\
= \ & E\left[\tilde{W}_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\left\{E\left[\frac{1}{h^{L+1}}\frac{D^*_{\boldsymbol{\pi}_s,\tau}}{f_{vx,\boldsymbol{\pi}_s}}K_{vx,h}\left(v_{\boldsymbol{\pi}_3}-v_{\boldsymbol{\pi}_s}, X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_s}\right)\mid\Omega_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\right]-E\left[D^*_{\boldsymbol{\pi}_s,\tau}\mid\omega_{\boldsymbol{\pi}_s}\right]\right\}\mid\zeta_{\boldsymbol{\pi}_1}\right] \\
= \ & O\left(h^{\overline{M}}\right)
\end{aligned}
$$

since the expectation

$$
\begin{aligned}
& E\left[\frac{1}{h^{L+1}}\frac{D^*_{\boldsymbol{\pi}_s,\tau}}{f_{vx,\boldsymbol{\pi}_s}}K_{vx,h}\left(v_{\boldsymbol{\pi}_3}-v_{\boldsymbol{\pi}_s}, X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_s}\right)\mid\Omega_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_2}\right] \\
= \ & \int\frac{1}{h^{L+1}}E\left[\frac{D^*_{\boldsymbol{\pi}_s,\tau}}{f_{vx,\boldsymbol{\pi}_s}}\mid\omega_{\boldsymbol{\pi}_s}\right]K_{vx,h}\left(v_{\boldsymbol{\pi}_3}-v_{\boldsymbol{\pi}_s}, X_{\boldsymbol{\pi}_3}-X_{\boldsymbol{\pi}_s}\right)f_{vx}\left(v_{\boldsymbol{\pi}_3}, X_{\boldsymbol{\pi}_3}\right)dv_{\boldsymbol{\pi}_3}dX_{\boldsymbol{\pi}_3} \\
= \ & E\left[D^*_{\boldsymbol{\pi}_s,\tau}\mid\omega_{\boldsymbol{\pi}_s}\right]+o\left(h^{\overline{M}}\right),
\end{aligned}
$$

where the second equality follows from Assumptions 3.1.1, 3.1.2, and properties of the bias-reducing kernel, Assumption 4.1.5.

Similarly, for a given $\boldsymbol{\pi}_s \in \{(\boldsymbol{\pi}_3(1), \boldsymbol{\pi}_2(2)), (\boldsymbol{\pi}_2(1), \boldsymbol{\pi}_3(2)), \boldsymbol{\pi}_2\}$, it follows from Assumptions 3.1.1, 3.1.2, 4.1.3, and 4.1.5, that

$$
\begin{aligned}
& E\left[\frac{1}{h^{L+1}}\left(\frac{D^*_{\boldsymbol{\pi}_s,\tau}}{f_{vx,\boldsymbol{\pi}_s}}\tilde{W}_{\boldsymbol{\pi}_3,\boldsymbol{\pi}_2}K_{vx,h}\left(v_{\boldsymbol{\pi}_1}-v_{\boldsymbol{\pi}_s}, X_{\boldsymbol{\pi}_1}-X_{\boldsymbol{\pi}_s}\right)\right)\mid\zeta_{\boldsymbol{\pi}_1}\right]-\Xi\left[v_{\boldsymbol{\pi}_1}, X_{\boldsymbol{\pi}_1}\right] \\
= \ & E\left[\frac{1}{h^{L+1}}E\left[\left(\frac{D^*_{\boldsymbol{\pi}_s,\tau}}{f_{vx,\boldsymbol{\pi}_s}}\chi_{\boldsymbol{\pi}_s}\right)\mid v_{\boldsymbol{\pi}_s}, X_{\boldsymbol{\pi}_s}\right]K_{vx,h}\left(v_{\boldsymbol{\pi}_1}-v_{\boldsymbol{\pi}_s}, X_{\boldsymbol{\pi}_1}-X_{\boldsymbol{\pi}_s}\right)\mid\zeta_{\boldsymbol{\pi}_1}\right]-\Xi\left[v_{\boldsymbol{\pi}_1}, X_{\boldsymbol{\pi}_1}\right] \\
= \ & \int\left\{\Xi\left(v_{\boldsymbol{\pi}_1}+h\boldsymbol{\nu}_{\mathbf{1}}, X_{\boldsymbol{\pi}_1}+h\boldsymbol{\nu}_{\mathbf{2}}\right)-\Xi\left(v_{\boldsymbol{\pi}_1}, X_{\boldsymbol{\pi}_1}\right)\right\}K_{vx}\left(\boldsymbol{\nu}\right)d\boldsymbol{\nu} \\
= \ & O\left(h^{\overline{M}}\right).
\end{aligned}
$$

Using the previous results it follows that

$$
E\left[p_{\boldsymbol{\pi}_s,\boldsymbol{\pi}_3}\left(\overline{\sigma}\right)\mid\zeta_{\boldsymbol{\pi}_1}\right] \quad = \quad E\left[D^*_{\boldsymbol{\pi}_1,\tau}\chi_{\boldsymbol{\pi}_1}\mid X_{\boldsymbol{\pi}_1}\right]-E\left[D^*_{\boldsymbol{\pi}_1,\tau}\chi_{\boldsymbol{\pi}_1}\right],
$$

and thus,

$$
\begin{aligned}
& E\left[p_{\boldsymbol{\pi}_1,\boldsymbol{\pi}_3}\left(\overline{\sigma}\right)-p_{\boldsymbol{\pi}_1(1)\boldsymbol{\pi}_2(2),\boldsymbol{\pi}_3}\left(\overline{\sigma}\right)-p_{\boldsymbol{\pi}_2(1)\boldsymbol{\pi}_1(2),\boldsymbol{\pi}_3}\left(\overline{\sigma}\right)+p_{\boldsymbol{\pi}_2,\boldsymbol{\pi}_3}\left(\overline{\sigma}\right)\mid\zeta_{\boldsymbol{\pi}_1}\right] \\
= \ & \left\{D^*_{\boldsymbol{\pi}_1}-E\left[D^*_{\boldsymbol{\pi}_1}\mid\omega_{\boldsymbol{\pi}_1}\right]\right\}I_{\tau,\boldsymbol{\pi}_1}\chi_{\boldsymbol{\pi}_1}+o(1)
\end{aligned}
$$

It follows then that the Hájek projection is given by

$$
V^*_{3,n\tau} \quad = \quad \frac{1}{n(n-1)}\sum_{i_1=1}^{n}\sum_{j_1\neq i_1}\xi_{i_1 j_1,\tau}+o(1)
$$

with

$$\begin{aligned}
\xi_{i_1 j_1,\tau} &= \left\{D^*_{i_1 j_1} - E\left[D^*_{i_1 j_1} \mid \omega_{i_1 j_1}\right]\right\} I_{\tau,i_1 j_1} \overline{\chi}_{i_1 j_1} \\
\overline{\chi}_{i_1 j_1} &= \left\{\frac{1}{(n-2)(n-3)} \sum_{i_2 \neq i_1, j_1} \sum_{j_2 \neq i_1, j_1, i_2} E\left[\tilde{W}_{\sigma\{i_1,i_2;j_1,j_2\}} \mid X_{i_1}, X_{j_1}\right]\right\}.
\end{aligned}$$

If follows from a Law of Iterated Expectations that

$$E\left[V^*_{3,n\tau}\right] = E\left[\xi_{i_1 j_1,\tau}\right] = 0.$$

Step 2. Variance of Hájek Projection

As in the proof of Lemma B.4, the variance of $V^*_{3,n\tau}$ is given by

$$\begin{aligned}
Var\left(V^*_{3,n\tau}\right) &= \left\{\frac{1}{n(n-1)}\right\}^2 \left\{\sum_{i_1=1}^{n} \sum_{j_1 \neq i_1} E\left[\xi_{i_1 j_1,\tau} \xi'_{i'_1 j'_1,\tau}\right]\right\} \\
&= \left\{\frac{1}{n(n-1)}\right\}^2 \left\{\sum_{i_1=1}^{n} \sum_{j_1 \neq i_1} \Lambda^*_{i_1,j_1}\right\}
\end{aligned}$$

where

$$\Lambda^*_{i_1,j_1} = E\left[\left\{E\left[D^*_{i_1 j_1} D^*_{i_1 j_1} \mid \omega_{i_1 j_1}\right] - E\left[D^*_{i_1 j_1} \mid \omega_{i_1 j_1}\right]^2\right\} I^2_{\tau,i_1 j_1} \overline{\chi}_{i_1 j_1} \overline{\chi}'_{i_1 j_1}\right].$$

Define

$$\Upsilon_n = n(n-1) Var\left(V^*_{1,n\tau}\right) = \frac{1}{n(n-1)} \left\{\sum_{i_1=1}^{n} \sum_{j_1 \neq i_1} \Lambda^*_{i_1,j_1}\right\}.$$

Step 3. Variance of $S_{3,n\tau}$

Given two different 6-tuples $\overline{\sigma}\{i_1, i_2; j_1, j_2; l_1, l_2\}$ and $\overline{\sigma}'\{i'_1, i'_2; j'_1, j'_2; l'_1, l'_2\}$, let

$$\Delta_{c,n} = Cov\left(s_{3,n}\left(\sigma\{i_1, i_2; j_1, j_2; l_1, l_2\}\right), s_{3,n}\left(\sigma'\{i'_1, i'_2; j'_1, j'_2; l'_1, l'_2\}\right)\right)$$

denote the covariance between $s_{3,n}(\overline{\sigma})$ and $s_{3,n}(\overline{\sigma}')$ when $\overline{\sigma}$ and $\overline{\sigma}'$ have $c = 0, 1, 2, 3, 4, 5, 6$ indices in common.

It follows from conditionally independence formation of links, implied by Assumption 3.1.2, and the conditional mean zero, $E\left[s_{3,n}\left(\sigma\{i_1, i_2; j_1, j_2; l_1, l_2\}\right) \mid \Omega_\sigma\right] = 0$, that $\Delta_{0,n} = \Delta_{1,n} = 0$.

Consider

$$
\begin{aligned}
\Delta_{2,n} &= E\left[s_{3,n}(\overline{\sigma}\{i_1,i_2;j_1,j_2;l_1,l_2\})s_{3,n}(\overline{\sigma'}\{i_1,i_2';j_1,j_2';l_1,l_2'\})'\right] \\
&= E\left[s_{i_1j_1}\left(\overline{\sigma}\right)s_{i_1j_1}\left(\overline{\sigma'}\right)'\right]+o(1) \\
&= E\left[\left\{E\left[\tilde{D}^*_{i_1j_1,\tau}\tilde{D}^*_{i_1j_1,\tau}\mid\omega_{i_1j_1}\right]-E\left[\tilde{D}^*_{i_1j_1,\tau}\mid\omega_{i_1j_1}\right]^2\right\}I^2_{\tau,i_1j_1}\tilde{W}_\sigma\tilde{W}_{\sigma'}\right]+o(1).
\end{aligned}
$$

Therefore, the variance of $Var(S^\dagger_{2,n\tau})$ can be expressed as

$$
\begin{aligned}
&\left(\frac{1}{\overline{m}_n}\right)^2\sum_{\overline{\sigma}}\sum_{\overline{\sigma}'}E\left[\left(s_{3,n}(\overline{\sigma})s_{3,n}(\overline{\sigma'})'\right)\right] \\
&+\left(4!\binom{n}{4}\right)^{-2}\sum_{i_1=1}^{n}\sum_{j_1\neq i_1}\left\{\sum_{k_1\neq i_1,j_1}\sum_{k_2\neq i_1,j_1,k_1}\sum_{l_1\neq i_1,j_1}\sum_{l_2\neq i_1,j_1,l_1}\Delta_{2,n}\right\} \\
&+O\left(\frac{1}{n^3}\right)\Delta_{3,n}+O\left(\frac{1}{n^4}\right)\Delta_{4,n}+O\left(\frac{1}{n^5}\right)\Delta_{5,n}+O\left(\frac{1}{n^6}\right)\Delta_{6,n}
\end{aligned}
$$

Notice that the term inside the brackets scaled by $((n-2)(n-3))^{-2}$ can be written as

$$
\begin{aligned}
&\left(\frac{1}{(n-2)(n-3)}\right)^2\sum_{k_1\neq i_1,j_1}\sum_{k_2\neq i_1,j_1,k_1}\sum_{l_1\neq i_1,j_1}\sum_{l_2\neq i_1,j_1,l_1}\Delta_{2,n} \\
=\;& E\left[\left\{E\left[D^*_{i_1j_1}D^*_{i_1j_1}\mid\omega_{i_1j_1}\right]-E\left[D^*_{i_1j_1}\mid\omega_{i_1j_1}\right]^2\right\}I^2_{\tau,i_1j_1}\overline{\chi}_{i_1j_1}\overline{\chi}'_{i_1j_1}\right] \\
=\;& \Lambda^*_{i_1,j_1}.
\end{aligned}
$$

As a result,

$$
Var\left[S^\dagger_{3,n\tau}\right] = \left(\frac{1}{n(n-1)}\right)^2\left\{\sum_{i_1=1}^{n}\sum_{j_1\neq i_1}\Lambda^*_{i_1,j_1}\right\}+o(1),
$$

and $Var\left[V^*_{3,n\tau}\right]-Var\left[S^\dagger_{3,n\tau}\right]=o_p(1)$.

The asymptotic equivalence results follows from similar arguments as in the proof of Lemma B.4. The proof is complete.

□